

(51) International Patent Classification ⁶ : G11C 11/00, 11/34, H01L 29/788		A1	(11) International Publication Number: WO 97/32309
		(43) International Publication Date: 4 September 1997 (04.09.97)	
(21) International Application Number: PCT/US97/01388 (22) International Filing Date: 29 January 1997 (29.01.97) (30) Priority Data: 08/607,951 28 February 1996 (28.02.96) US (71) Applicant: SANDISK CORPORATION [US/US]; 140 Caspian Court, Sunnyvale, CA 94089 (US). (72) Inventors: GUTERMAN, Daniel, C.; 305 Jacaranda Drive, Fremont, CA 94539 (US). SAMACHISA, Gheorghe; 6858 Castlerock Drive, San Jose, CA 95120 (US). FONG, Yupin, K.; 2100 Springwater Drive, Fremont, CA 94539 (US). HARARI, Eliyahou; 104 Auzeais Court, Los Gatos, CA 95030 (US). (74) Agents: CASERZA, Steven, F. et al.; Flehr, Hohbach, Test, Albritton & Herbert L.L.P., Suite 3400, 4 Embarcadero Center, San Francisco, CA 94111-4187 (US).		(81) Designated State: JP. Published <i>With international search report.</i>	
(54) Title: EEPROM WITH SPLIT GATE SOURCE SIDE INJECTION			
(57) Abstract A semiconductor EPROM cell (101) with memory arrays organized in sectors with each sector formed of a single column or group of columns with their control gates (108) connected in common.			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgyzstan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	KZ	Kazakhstan	SG	Singapore
CH	Switzerland	LJ	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovakia
CM	Cameroon	LR	Liberia	SN	Senegal
CN	China	LT	Lithuania	SZ	Swaziland
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	LV	Latvia	TG	Togo
DE	Germany	MC	Monaco	TJ	Tajikistan
DK	Denmark	MD	Republic of Moldova	TT	Trinidad and Tobago
EE	Estonia	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	UG	Uganda
FI	Finland	MN	Mongolia	US	United States of America
FR	France	MR	Mauritania	UZ	Uzbekistan
GA	Gabon			VN	Viet Nam

EEPROM WITH SPLIT GATE SOURCE SIDE INJECTION

INTRODUCTIONRelated Applications

This application is a continuation-in-part of U.S. Serial Number 08/193,707 filed February 9, 1994, which in turn is a divisional of U.S. Serial Number 07/820,364, filed January 14, 1992, now U.S. Patent 5,313,421 issued May 17, 1994.

Technical Field

This invention pertains to semiconductor memory cells and arrays, more particularly to electrically erasable programmable read only memories.

Background

Erasable programmable read only memories (EPROMs) and electrically erasable programmable read only (EEPROMs) are well known in the art. These devices have the ability to store data in non-volatile fashion, while also being capable of being erased and rewritten as desired. EPROM devices are typically erased by exposing the integrated circuit device to ultraviolet radiation, while EEPROMs allow erasure to be performed electrically.

One form of EEPROM device includes a so-called "split-gate" electrode, in which the control gate includes a first portion overlaying a floating gate and a second portion directly overlaying the channel. Such a split gate structure is described in a 5-Volt-Only Fast-Programmable Flash EEPROM Cell with a Double Polysilicon Split-Gate Structure by J. Van Houdt et al, Eleventh IEEE Non-Volatile Semiconductor Workshop, February 1991, in which charge is injected into the floating gate from the source side of the cell. U.S. Patent 4,652,897 describes an EEPROM device which does not utilize a split-gate,

1 but which also provides injection to the floating gate from the
2 source side of the device.

3 As described in the above referenced U.S. Patent
4 4,652,897, memory cells are typically arranged in an array, as
5 is well known in the art. One form of such an array utilizes
6 buried diffusions, in which source and array regions are
7 covered with a fairly thick layer of insulating material. This
8 is shown for example, in U.S. Patents 4,151,020; 4,151,021;
9 4,184,207; and 4,271,421. Such buried diffusion devices often
10 utilize a virtual ground approach, in which columns connecting
11 the sources of a first column of memory cells also serves to
12 connect drains of an adjacent column of memory cells.

13 While many EEPROM devices utilize two layers of
14 polycrystalline silicon, one for the formation of the floating
15 gate, and the other for the formation of the control gate and
16 possibly electrical interconnects, other EEPROM devices utilize
17 three layers of polycrystalline silicon. For example, U.S.
18 Patent 4,302,766 provides a first polycrystalline silicon layer
19 for the floating gate, a second polycrystalline silicon layer
20 for the control gate, and a third polycrystalline silicon layer
21 coupled through an erase window to a portion of the first
22 polycrystalline silicon layer for use during erasure of the
23 cell. U.S. Patent 4,331,968 also uses a third layer of
24 polycrystalline silicon to form an erase gate, while U.S.
25 Patent 4,462,090 forms an addressing gate electrode utilizing
26 a third layer of polycrystalline silicon. U.S. Patent
27 4,561,004 and 4,803,529 also use three layers of
28 polycrystalline silicon in their own specific configurations.

29 Japanese Patent Publication 61-181168 appears to utilize
30 three layers of polycrystalline silicon to provide additional
31 capacitive coupling to the floating gate. Japanese Patent
32 Publication 63-265391 appears to pertain to a buried diffusion
33 array, possibly utilizing virtual grounds.

34 European Patent Application 0373830 describes an EEPROM in
35 which two polycrystalline silicon layers are used, with the
36 second layer of polycrystalline silicon having two pieces, one
37 of which provides the erase function, and one of which provides
38 the steering function.

1 "A New Flash-Erase EEPROM Cell With a Sidewall Select-Gate
2 on its Source Side" by K. Naruke et al. IEDM-89-603 and U.S.
3 Patent 4,794,565 describe an EEPROM utilizing a side wall
4 select gate located on the source side of the field effect
5 transistor.

6 "EPROM Cell With High Gate Injection Efficiency" by
7 M. Kamiya et al. IEDM 82-741, and U.S. Patent 4,622,656
8 describe an EEPROM device in which a reduced programming
9 voltage is provided by having a highly doped channel region
10 under the select gate, and the channel region under the
11 floating gate being either lightly doped or doped to the
12 opposite conductivity type, thereby providing a significant
13 surface potential gap at the transition location of the
14 channel.

15 In recent years there has been significant interest in
16 producing high capacity FLASH memory devices which use split-
17 gate, source-side hot electron programming, in place of the
18 more conventional drain-side channel hot electron (CHE)
19 mechanism.

20 The reasons for this include its inherently lower write
21 power requirement (1/10th that of CHE or less), facilitating
22 low voltage operation and higher write speeds via increased
23 parallelism. In addition, the split gate structure is not
24 susceptible to "overerase" related problems (a problem for
25 single gate FLASH memories such as ETOX), and does not
26 experience programming difficulty due to strong overerase,
27 which can hinder programming after an erasure operation in
28 split-gate CHE programming devices.

29 In view of these benefits, SunDisk Corporation has
30 patented FLASH memory cell and array variants which use source
31 side injection integrated with SunDisk's proprietary thick
32 oxide, poly-to-poly erase tunneling technology, to make a
33 highly scalable, reliable, low power programming cell (D.C.
34 Guterman, G. Samachiasa, Y. Fong and E. Harari, U.S. Patent No.
35 5,313,421).

36 The concept of a multi-bit storage non-volatile cell using
37 a split gate structure was described by G.S. Alberts and H.N.
38 Kotecha (Multi-bit storage FET EAROM cell, IBM Technical

1 Disclosure Bulletin, Vol. 24 No. 7A, p. 3311, Dec. 1981). They
2 describe a two-poly, three transistor element-in-series cell,
3 in which the center transistor's channel is controlled directly
4 by the poly2 control gate (which also serves as the cell select
5 gate), and each of the two end transistor channels are
6 controlled by corresponding poly1 floating gates, which in turn
7 are capacitively coupled to the control gate, thereby realizing
8 a plurality of bits in the one physical cell structure.

9 Recently, at the 1994 IEDM, Bright Microelectronics along
10 with Hyundai presented a similar dual-bit split-gate cell,
11 integrated into a contactless, virtual ground array, and using
12 source side injection programming (Y.Y. Ma and K. Chang, U.S.
13 Patent No. 5,278,439 - referred to henceforth as the Ma
14 approach). One structural difference here from the IBM
15 approach is their separation of the capacitively coupling
16 control gates, which are formed in poly2, and the select gate,
17 which is formed in poly3.

18 In the Ma approach, they use "conventional" negative
19 control gate driven tunneling through an ultra-thin poly1 gate
20 oxide (about 100Å or less). This erase approach poses some
21 serious limitations. Erase of one of the two storage
22 transistors uses floating gate to drain tunneling through the
23 ultra-thin oxide, accomplished by biasing the drain to 7v and
24 corresponding control gate to -10v. Because both of these
25 lines run perpendicular to the select gate, this forces a block
26 of cells which are to be simultaneously erased (e.g. a sector)
27 to be bit line oriented, as opposed to the more conventional
28 word line (select gate) oriented block; i.e. its sector must be
29 column organized and thus it cannot be row organized. (For
30 example, a sector could be two columns of floating gates
31 straddling a bit line/diffusion, including the right hand
32 floating gates of the left side cells' floating gate pair plus
33 the left hand floating gates of the right side cells.)
34 This leads to the following disadvantages in the Ma
35 implementation:

36 (1) Limited to column sector architecture; i.e. cannot readily
37 support the higher read performance row oriented sector
38 architecture. (Since here, within a sector, both erase anode

1 and corresponding control gates run perpendicular to row line
2 direction, this precludes the massively parallel "chunk"
3 implementation of the row oriented sector, which can
4 simultaneously access large numbers of cells within that
5 sector).

6 (2) Requires ultra-thin, approximately 100Å, tunneling oxide,
7 imposing following limitations:

8 * Scaling limitation associated with pushing the limits
9 of usable oxide thicknesses, plus the additional area needs
10 associated with maintaining adequate coupling requirements,
11 which must combat the inherently high capacitance per unit area
12 of such a thin oxide;

13 * A myriad of potential retention/reliability problems
14 inherent to using ultra-thin oxide, combined with the parasitic
15 band-to-band tunneling/hole injection associated with the high
16 substrate fields adjacent to the diffusion anode; and

17 * Negative gate bias requirements on control gate, to
18 limit band-to-band injection problems, impose process and
19 circuit complexity, plus potentially more layout area
20 requirement.

21

22

SUMMARY OF THE INVENTION

23 In accordance with the teachings of this invention, novel
24 memory cells are described utilizing source-side injection.
25 Source-side injection allows programming utilizing very small
26 programming currents. If desired, in accordance with the
27 teachings of this invention, to-be-programmed cells along a
28 column are programmed simultaneously which, due to the small
29 programming current required for each cell, does not require an
30 unacceptably large programming current for any given
31 programming operation. In one embodiment of this invention,
32 the memory arrays are organized in sectors with each sector
33 being formed of a single column or a group of columns having
34 their control gates connected in common. In one embodiment, a
35 high speed shift register is used in place of a row decoder in
36 order to serially shift in the data for the word lines, with
37 all of the data for each word line of a sector being contained
38 in the shift register on completion of its serial loading. In

1 one embodiment, additional speed is achieved by utilizing a
2 parallel load d buffer register which receives data in parallel
3 from the high speed shift register and holds that data during
4 the write operation, allowing the shift register to receive
5 serial loaded data during the write operation for use in a
6 subsequent write operation. In one embodiment, a verification
7 is performed in parallel on all to-be-programmed cells in a
8 column and the bit line current monitored. If all of the to-
9 be-programmed cells have been properly programmed, the bit line
10 current will be substantially zero. If bit line current is
11 detected, another write operation is performed on all cells of
12 the sector, and another verify operation is performed. This
13 write/verify procedure is repeated until verification is
14 successful, as detected by substantially zero bit line current.

15 Among the objectives of the novel cells constructed in
16 accordance with this invention are avoidance of programming
17 limitations such as:

- 18
- 19 1. High Channel Currents (Power) required for
20 Programming.
 - 21 2. High Drain Voltage Requirements, which increase with
22 increased levels of erasure.
 - 23 3. Loss of Read Performance associated with an increase
24 in Programming Efficiency via Heavy Channel doping.
 - 25 4. Program Wearout Associated with Maintaining a High
26 Drain Bias on Cells exposed to this bias, including
27 both those cells targeted for programming and those
28 cells not targeted but still exposed to the voltage.
- 29

30 In an alternative embodiment of this invention, a multi-
31 bit memory cell is taught utilizing a 3-poly, 3 transistor
32 element-in-series cell in which the center transistor's channel
33 is controlled directly by the poly 3 control gate (which serves
34 as both a cell select gate and erase anode) and each of the two
35 end transistor channels are controlled by corresponding poly1
36 floating gates, which in turn are capacitively coupled to the
37 poly 2 control or steering gates, thereby realizing a plurality
38 of bits in the one physical cell structure.

1 The multi-bit cell contains two bits per unit memory cell,
2 coming from two floating gate portions, each having their own
3 control gate (which, in the virtual ground array, runs parallel
4 to the bit lines), and sharing one select gate, placed
5 physically between them (which, in the virtual ground array,
6 runs perpendicular to the bit lines). The diffusion BN+
7 source/drains straddle the two floating gates, on their
8 opposite facing channel edges to those adjacent the select
9 gate/transfer channel.

10 Unlike a single floating gate cell, because here the two
11 floating channels lie in a series configuration, the programmed
12 threshold voltage level of each floating gate must be limited
13 in its upper value in order to be readable (similarly to the
14 Toshiba NAND cell). In this way, either floating gate channel
15 can be unconditionally turned on (i.e. independent of its
16 stored state) when appropriate bias is applied to its
17 corresponding control gate, when reading the state of the other
18 floating gate.

19

20 BRIEF DESCRIPTION OF THE DRAWINGS

21 Figures 1a, 1b, and 1c, are cell layout, cross-sectional
22 diagram, and equivalent circuit schematic of one embodiment of
23 this invention;

24 Figure 1d is a plan view of one embodiment of an array
25 consisting of a plurality of cells of Figures 1a-1c;

26 Figure 1e is a block diagram depicting a memory array
27 organized by sectors, with appropriate control circuitry;

28 Figure 1f depicts the operation of one embodiment of a
29 memory array organized by sectors as shown in Figure 1e;

30 Figure 1g is a plan view depicting an alternative array
31 embodiment utilizing cells depicted in Figures 1a-1c;

32 Figure 2a is a cross-sectional view depicting an
33 alternative embodiment of this invention similar that of Figure
34 1b;

35 Figure 2b is a plan view of one embodiment of an array of
36 memory cells constructed utilizing cells depicted in the cross-
37 sectional view of Figure 2a;

1 Figure 2c is a diagram depicting the organization and
2 operating condition of an array such as that of Figure 2b;

3 Figure 3 is a graph depicting the operation of a memory
4 cell of Figure 1b;

5 Figure 4 depicts the electrical field distribution along
6 channels of the device of Figure 5;

7 Figure 5 is a cross-sectional view one embodiment of a
8 2-poly cell of this invention;

9 Figure 6 is a cross-sectional view of another embodiment
10 of a 2-poly cell of this invention;

11 Figure 7a is a plan view depicting a portion of a process
12 sequence utilized in accordance with one embodiment of this
13 invention;

14 Figure 7b is a cross-sectional view of the embodiment
15 shown in the plan view of Figure 7a;

16 Figure 8 is a cross-sectional view depicting a fabrication
17 step suitable for use in accordance with the teachings of this
18 invention;

19 Figures 9a and 9b are top and cross-sectional views,
20 respectively of one embodiment of a multiple-bit memory cell
21 structure of this invention;

22 Figure 10a is a schematic diagram of one multi-bit cell of
23 this invention;

24 Figure 10b is a circuit diagram depicting one embodiment
25 of an array of multiple-bit memory cells of this invention,
26 such as those of Figures 9A and 9B;

27 Figure 10c is a circuit diagram depicting one embodiment
28 of an array of cells as shown in Figure 10b plus segment decode
29 transistor matrix for both bit lines and steering lines;

30 Figures 11a through 11e are detailed top and cross-
31 sectional views; and

32 Figures 12a-12f are cross sectional views depicting
33 fabrication steps suitable for use in fabricating multi-bit
34 memory cells in accordance with this invention.

35

36 DESCRIPTION OF SPECIFIC EMBODIMENTS

37 The cell layout, cross-sectional diagram and equivalent
38 circuit schematic of one embodiment are shown in Figures 1a,

1 lb, and 1c, respectively. Similar reference numerals have been
2 used in Figures 1a, 1b, and 1c. Referring to the cross-
3 sectional view of Figure 1b, this embodiment of the novel
4 EEPROM cell 101, of this invention includes a buried source
5 region 102 and a buried drain region 103, each being buried by
6 a relatively thick layer of dielectric 104 and 105,
7 respectively. Channel region 106 is divided into two portions,
8 a first portion 106-1 which is influenced by the third layer
9 polycrystalline silicon 109 and which forms a select gate, and
10 a second portion 106-2 which is influenced by floating gate 107
11 formed of a first layer of polycrystalline silicon and which,
12 in turn, is influenced by control gate 108 formed of a second
13 layer polycrystalline silicon. As is well known in the art,
14 suitable dielectric layers such as thermally grown oxide are
15 located between channel 106 and polycrystalline silicon layer
16 109 and polycrystalline silicon layer 107. Similarly, suitable
17 dielectric layers such as oxide or composite oxide/nitride are
18 formed between the three layers of polycrystalline silicon.
19 Polycrystalline metal silicide can be used in place of one or
20 more of the polycrystalline silicon layers 108 and 109. If
21 desired, a highly-doped P+ region 120 is used within channel
22 106-2 adjacent buried drain region 103. This region 120 is
23 formed, for example, as a double diffused MOS (DMOS) region in
24 order to establish the threshold voltage V_t of the memory
25 transistor including channel 106-2. This helps to provide a
26 stable threshold voltage, even though the amount of charges
27 trapped in the gate oxide layer in the vicinity of the gap
28 between 106-1 and 106-2 tends to increase with a large number
29 of programming cycles.

30 An example of operating conditions and levels associated
31 with the embodiment of Fig. 1b are shown in Table 1. High
32 efficiency programming comes about by the ability to
33 simultaneously create a high field region in channel 106-2
34 under the floating gate, which under the bias conditions of
35 Table 1 occur near the gap between channels 106-1 and 106-2
36 (see above mentioned IEDM article of Kamiya for theory) while
37 maintaining a low channel/current. Since this high field
38 causes electron injection to floating gate 107 near the source

1 side of channel 106-2, this type of operation is termed
2 "source-side" injection. This mechanism provides high
3 efficiency, low power programming by maintaining a low channel
4 current via word line 109 throttling by using a bias operating
5 near channel threshold, $V_{T_{p3}}$. A major attribute of this type
6 of operation is that it allows for a high drive condition in
7 floating gate channel 106-2 under the floating gate (in fact it
8 thrives on it), offering high-performance read, without
9 degrading programming performance. This is because the very
10 weak drive condition on the select transistor of channel 106-1
11 is established via the throttling mentioned above to achieve
12 the high fields in the vicinity of the poly 3/poly 1 gap.
13 These fields accelerate the electrons to sufficiently energetic
14 levels (i.e. > 3.1 eV) to surmount the Si/SiO₂ interface
15 barrier at the source side of floating gate 107. Furthermore,
16 there is a significant vertical component to that field (i.e.
17 normal to the Si/SiO₂ surface) driving the electrons up to the
18 surface of channel 106, and thereby assisting the injection
19 into floating gate 107. No read performance penalty is
20 incurred to establish this high field condition. This is in
21 stark contrast to conventional drain side programming, wherein
22 efficient program requires strong channel saturation which
23 shuns high floating gate channel drives, strong overerase, or
24 a weakly turned on series select transistor. These problems
25 with drain side programming dictate high channel currents, care
26 in overerase, potentially high drain voltages, and unfavorable
27 fields (potentially subducting the channel below the surface at
28 the drain side and driving electrons downward away from the
29 floating gate).

30 Consequently, in accordance with the teachings of this
31 invention, programming efficiencies (I_G/I_D) ranging from
32 10^{-5} to 10^{-3} are possible, with I_D in the range of 1mA during
33 programming, which is two to three orders of magnitude smaller
34 than conventional drain side programming. This offers the
35 potential for very fast system level programming by allowing
36 the programming of 100 times as many memory cells in parallel,
37 thereby achieving a 100 fold increase in effective program
38 speed compared with prior art drain side programming.

TABLE 1

State Table & Operating Conditions (Fig. 1b)

Node Operation	Poly 3 (Word line)	Poly 2 (Steering Gate)	Drain (BN & Drain)	Source & (BN Source)
R STANDBY	0v	0v	1.0v or	1.0v or
E			0v	0v
R L READ SELECTED	5v	0v	1.0v or	0v or
E A			0v	1.0v
A T READ UNSELECTED	5v	0v	1.0v	1.0v
D E				
D				
E R ERASE UNSELECTED	5v	0v	0v	0v
R E				
A L				
S A				
E T ERASE Option 1	5v	-10to-17v	0v	0v
E or Option 2	12-22v	0v	0v	0v
D				
P R PROGRAM	H1.5v	14-20v	5-7v	0v
R E SELECTED				
O L				
G A				
R T PROGRAM	0v	14-20v	5-7v	0v
A E UNSELECTED	H1.5V	14-20v	5-7v	5-7v
M D	0v	14-20v	0v	0v

A major feature of the cell of this invention is the decoupling of the select function (in this case poly 3 select transistor 110 in Figure 1b) from the steering function (poly 2 control gate 108). During programming, this allows the independent control of cell selection/drain current throttling via poly 3 word line 109 bias (biased at slightly higher than $V_{T_{p3}}$) and strong positive voltage coupling onto floating gate 107 (by raising poly 2 control gate 108 to a high voltage, such as about 12 volts). Also, in accordance with the teachings of this invention, the drain voltage can be adjusted independently of steering and select transistor voltage levels, to optimize programming.

1 During read, the decoupling feature of this invention
2 provides two important advantages, and one exciting side
3 benefit.

4
5 1. The ability to set control gate 108 at the optimum voltage
6 level for memory state sensing, i.e. the best balanced
7 reference point for both programmed and erased states. This
8 independence is in contrast to conventional cells wherein the
9 control gate also serves as the select transistor, dictating a
10 voltage level consistent with selection (e.g. $V_{cc} = 5v \pm 10\%$).

11
12 2. Improved margin by virtue of being a fixed, (potentially
13 regulated) reference voltage, eliminating the V_{cc} variation of
14 $\pm 10\%$ inherent to the word line bias levels. (This alone
15 could improve the floating gate memory window by about 0.6v).

16
17 3. A side benefit of the ability to independently set the
18 control gate voltage bias discussed above, offers the
19 possibility of a simple way for re-referencing the memory cell
20 for multi-state (i.e. more than conventional 2-state) encoded
21 data. For example if the cell is encoded into three level
22 states, (such as logical 1 = strongly erased/high conducting,
23 logical 2 = partially programmed/ weakly conducting; logical 3
24 = strongly programmed,) then the control gate voltage can be
25 set at two different levels in a two pass read scheme. For
26 example, in the first pass read the control gate voltage would
27 be set at about 0v to discriminate between the logical 1 state
28 and the logical 2/logical 3 states. In the second pass read
29 the control/gate voltage is set to about 2v, to discriminate
30 between the logical 3 state and the logical 1/logical 2 states.
31 By combining the information of this two pass read (e.g.
32 according to Table 2) the original state of the 3 state cell is
33 recovered. This biasing can be done independently of sense amp
34 reference cell considerations allowing a single sense
35 amp/reference cell circuit to detect the different states via
36 a multi-pass read scheme.

37

TABLE 2

READ	PASS 1	PASS 2
STATE	[Ref. = 0v]	[Ref. = 2]
1	Hi	Hi
2	Lo	Hi
3	Lo	Lo

The two options for erase operation/bias conditions shown in Table 1 stem from two different sets of considerations. The first option shown brings poly 2 control gate 108 to a large negative voltage, but allows poly 3 word line 109 to remain at a low voltage (e.g. 0v to 5v). This is desirable since the word lines and their decoders are preferably high performance, and repeated many times with a tightly pitched requirement, making high voltage word line requirements more difficult and real estate consuming to implement. Poly 2 control or steering gate 108 on the other hand could be common to a multiplicity of word lines (e.g. a sector consisting of 4 or more word lines), putting less demands on real estate and minimal impact to performance. Possible drawbacks of this approach are the process and device requirements to support both negative as well as positive polarity high voltage circuitry, and reduced steering effectiveness in that the channel cannot assist in steering by virtue of it being held at or near ground (i.e. can't go to large negative potential).

Note that poly 2 is used only as a steering electrode during all three operations. Poly 3, which is the word line connection to the X-decoder, only sees 0V to 5V (other than for erase option 2), and its capacitance can be made relatively small. It is relatively easy to generate +5V and -17V on poly 2 since both writing and erasing are slow operations relative to reading and there is no DC current drain. The -17V does require high voltage PMOS in the erase decode, but the +5V on poly 3 aids in reducing the maximum negative voltage required on poly 2 during erase.

The second option of using high word line voltage bias for erase eliminates both of the above potential drawbacks, but

1 burdens the high performance, tightly pitched word line/driver
2 with high voltage requirement.

3 Figure 1d is a plan view of one embodiment of an array
4 consisting of a plurality of cells constructed as just
5 described with respect to Figures 1a-1c, and using similar
6 reference numerals. Also shown, are channel stop isolation
7 regions 180.

8 Figure 1e shows a block diagram of a memory array similar
9 to that shown in the plan view of Figure 1d which is organized
10 by sectors, with appropriate control circuitry. Operation of
11 one embodiment of such a memory array organized by sectors is
12 shown in Figure 1f, where the abbreviations used have the
13 following meanings:

14 FLT = float

15 V_{BE} = bit line erase voltage

16 V_{WE} = word line erase voltage

17 DI = data in

18 DIV = data in during verify operation

19 V_{CEU} = control gate erase voltage - unselected

20 V_{CE} = control gate erase voltage - selected

21 S.A. = sense amplifier

22 V_{CH} = control gate margin voltage (during verify operation)

23 V_{CP} = control gate program voltage

24 V_{CR} = control gate read voltage

25 V_{CE} = control gate erase voltage

26

27 As shown in Figures 1e and 1f, in this embodiment sectors
28 are formed by a single column or a group of columns having
29 their control gate connected in common. This allows a high
30 speed shift register to be used in place of a row decoder in
31 order to serially shift in a whole block of column data for the
32 word lines, with the data for each word line being contained in
33 the shift register on completion of its serial loading. The
34 use of such a high speed shift register saves circuit area on
35 an integrated circuit by serving both encoding and latching
36 functions normally performed by a row decoder. Furthermore,
37 speed is improved by including a parallel loaded buffer
38 register which receives data in parallel from the high speed

1 shift register and holds that data during the write operation.
2 While the write operation takes place based upon the data
3 stored in the buffer register, the high speed serial shift
4 register receives the next block of data for subsequent
5 transfer to the buffer register for the next write operation.
6 In one embodiment of this invention, each sector has an
7 associated latch for tagging that sector in preparation for an
8 erase of a plurality of tagged sectors.

9 In one embodiment of this invention, a sector is formed in
10 a group of four cell columns, each column being 1024 bits tall
11 with a common control gate and an associated sector latch. In
12 this embodiment, verification of programming is performed in
13 parallel on all to-be-programmed cells in a single column.
14 Logical 0 state cells have word lines at 0 volts while logical
15 1 state cells have word lines at a positive voltage, such as 5
16 volts. The control gate and drain voltages are reduced to a
17 verify level to allow for proper margin testing and the bit
18 line current is monitored. If all of the to-be-programmed
19 cells have been properly programmed, the bit line current will
20 be 0 or substantially so. If not, it is known that one or more
21 of the to-be-programmed cells in the column have not been
22 properly programmed, and another write operation is performed
23 on the entire column, thereby assuring that any incompletely
24 ones of the to-be-written cells are again written. An
25 additional verify step is performed to verify that the column
26 has been properly programmed.

27 One embodiment of a process suitable for fabricating the
28 structure having the cross-sectional view of Figure 1b is now
29 described. This embodiment can be implemented in a very
30 small area with no need for an isoplanar oxide when utilizing
31 a virtual ground, allowing an isolation implant to be placed in
32 the remaining field which is not covered by diffusions or
33 polycrystalline silicon and avoids susceptibility to substrate
34 pitting associated with the SAMOS etch in the field isolation
35 region not covered by poly 1. This is achieved, for example,
36 with the following process sequence:

37

- 1 1. Form BN⁺ bit lines in vertical strips. Grow approximately
2 1500Å oxide on top of BN⁺, and approximately 200-300Å gate
3 oxide.
4
- 5 2. As shown in Figs. 7a and 7b, deposit poly 1 to a suitable
6 conductance and etch in horizontal strips perpendicular to the
7 BN⁺ diffusion. Fill the spaces between adjacent strips of poly
8 1 with deposited oxide, such as CVD followed by an etch back.
9 This approach protects the field isolation regions, and if
10 desired it can be preceded by a boron channel stop implant.
11
- 12 An alternative for steps 1 and 2 of the above process sequence
13 is forming horizontal strips of isolation oxide first, and then
14 depositing P₁ and etched back in RIE to fill and planarize the
15 horizontal grooves between adjacent strips of isolation oxide.
16
- 17 3. Form thin dielectric 140 such as ONO of approximately 300-
18 400 Å. covering poly 1 strips.
19
- 20 4. Deposit poly 2 and form a suitably thick dielectric
21 overlayer (e.g., approximately 2000-3000 Å of CVD densified
22 oxide). Etch this oxide and underlying poly 2 in long vertical
23 strips parallel to bit line (BN⁺) diffusions.
24
- 25 5. Form oxide spacers 62 along edges of poly 2 and use edge
26 of these spacers to define the floating gate by etching off
27 exposed poly 1 (i.e. poly 1 not covered by poly 2 or by
28 spacer).
29
- 30 6. Form tunnel erase oxide in a conventional manner, as
31 described in U.S. patent application serial number 323,779,
32 filed March 15, 1989, over exposed edges of poly 1 as well as
33 gate oxide over the channel of the select transistor (channel
34 106-1 in Figure 1b).
35
- 36 7. Deposit poly 3 or polysilicide, and form word lines in
37 horizontal strips.
38

Another embodiment for achieving a virtual ground cell without the use of the buried diffusion formed early in the process is now described. In place of the BN+ of step 1, after step 6 a photoresist (PR) masked arsenic source/drain implant 103a is used, self-aligned to one edge of poly 2 108 after poly 1 107 stack formation but leaving an unimplanted region along the other edge to become the poly 3 controlled select transistor channel (see Figure 8). The isolation oxide thickness formed earlier between poly 1 strips is made sufficiently thick to withstand the self-aligned poly 2/1 stack etch without exposing the substrate to pitting, but thin enough such that following this stack etch it is readily removed to expose the substrate to the source drain implant. This offers the benefit of reduced thermal drive of the arsenic junction laterally facilitating scaling. The remainder of the process steps of this embodiment follows the prior embodiment.

In summary, the novel cell of this invention offers the following benefits.

- * Very low programming current.
- * Low programming drain voltage requirement/eliminating the need for high voltage.
- * Immunity of Programmability to increased levels of erase.
- * Adjustability of memory state for optimum read of both program and erased states.
- * Improved margin by elimination of sensitivity to $\pm 10\%$ Vcc variation on the steering element.
- * Potential for pure low voltage word line/decoder implementation.
- * Facilitates multi-state cell sensing.
- * Reduced susceptibility to source side hot-electron programming induced trapping by establishing a separate threshold control region at the drain.

A second array embodiment is similar to that of Figure 1d but uses the cell embodiment shown in Figure 1b, to form a row

1 oriented sector architecture , is shown in Figure 1g. A sector
2 consists of a group of rows, four in this example, which are
3 erased together. Erase uses option 2 of Table 1, for this row
4 oriented sector architecture, bringing all the poly 3 word
5 lines of a sector to high voltage. The poly 2 steering gate is
6 common to a group of N sectors where N can range from 1 to the
7 full size of the memory array. Similarly the BN+ columns can
8 alternatively continuously span the full length of the array or
9 be broken down into a collection of shorter length, local
10 columns. These connect to a global (full array length) column
11 through a select transistor driven by an additional level of
12 decoding. The local columns can range from 1 to N sectors.
13 The preferred embodiment is to have local columns span the same
14 number of sectors as the poly 2 steering gate. A preferred
15 number of sectors, N, spanned by local columns and poly 2
16 steering is around 8. This is because if N is much smaller
17 than 8, the area overhead for local column section devices and
18 poly 2 steering gate routing is high in relation to the area of
19 arrayed cells, while if N is much larger than 8, the benefits
20 of having localized columns and poly 2 steering diminish.
21 These benefits are: (1) reduced bit line capacitance improving
22 read performance; (2) reduced repetitive exposure on unselected
23 sectors to the raised voltage conditions on drains and steering
24 electrodes when programming one sector within the N-sector
25 group, and associated potential disturb phenomena; and (3)
26 increased confinement of array related failures thereby
27 increasing the efficiency of replacing such failures. Read,
28 program and unselected conditions are as described in Table 1,
29 during read or program. The poly 3 word line in the selected
30 row within the selected sector is turned on, 5 volts for read
31 and approximately 1 volt for programming. Concurrently, the
32 drain to source bias conditions are applied to the columns,
33 approximately 5 volts for program and approximately 1.0-1.5
34 volts for read. In one embodiment, alternate bits in a
35 selected row are programmed simultaneously, thereby permitting
36 all bits in a selected row to be programmed utilizing two
37 programming operations. In a similar manner, in this
38 alternative embodiment, alternate bits in a selected row are

1 read (or verified) simultaneously, thereby permitting all bits
2 in a selected row to be read (or verified) utilizing two read
3 (or verify) operations. After one row in the sector has
4 finished reading or writing, the next row is selected, and so
5 forth to the end of the sector. The resulting row oriented
6 sector architecture and array operation is much more
7 conventional than the column oriented sector of the first
8 embodiment, and consequently operates in a more traditional
9 manner. Both embodiments share the intrinsic low power
10 capability of this invention, but the row oriented sector
11 embodiment requires, in addition, a full complement of data
12 registers to support massively parallel write and verify
13 features.

14 Figure 2a shows an alternative array embodiment of this
15 invention which does not utilize buried diffusion regions.
16 Thus, source region 102 and drain region 103 are formed in a
17 conventional manner and not buried by a thick dielectric layer
18 as is the case in the embodiment of Figure 1b. A plurality of
19 memory cells are shown in Figure 2a along a cross section of a
20 typical array structure, with elements of one such cell
21 numbered using reference numerals corresponding to similar
22 structure in Figure 1b. Table 3 depicts an example of the
23 operating conditions appropriate for the embodiment of Figure
24 2a. This a more traditional cell approach compared to the
25 buried diffusion cell, with source/drain diffusions formed
26 after all the polycrystalline silicon structures are formed.
27 It requires one drain contact to metal bit line for every 2
28 cells, making it approximately 30% to 50% larger than the
29 buried diffusion cell with similar layout rules. In all other
30 respects, this alternative embodiment offers the same benefits
31 as listed above for the buried diffusion embodiment of Figure
32 1b.

33 Figure 2b is a plan view of one embodiment of an array of
34 memory cells constructed as described above with reference to
35 Figure 2a.

36 Figure 2c is an equivalent circuit diagram depicting the
37 organization of such a memory array in sectors, with
38 appropriate operating conditions and voltages shown. The

1 preferred embodiment for a sector organized array uses two word
2 lines which straddle a source line as part of a sector, along
3 with their associated poly 2 steering gates and source line.
4 A full sector consists of some multiple of such pairing (e.g.
5 2 such pairs or 4 word lines, each word line containing 128
6 bytes and overhead cells, and straddling two source lines,
7 constitute one sector).

8 As shown in the embodiment of Figure 2c, the steering lines
9 are connected together within a sector as are the source lines
10 (i.e. a sector which consists of row lines grouped together
11 respectively and driven by common drivers). The embodiment
12 described here confines the write operation to the sector being
13 written to, while the bit line bias conditions (2.5v during
14 read and approximately 5v possible during write) are non-
15 disturbing to the cells because the bias is applied to the
16 select transistor side of the cell and not to the floating gate
17 side. In a two state cell, to write the cell to a logical one,
18 the bit line is held at zero volts, causing the cell to program
19 via source-side injection. Conversely, to inhibit writing, the
20 bit line is held high (typically about 5 volts), thereby
21 cutting off the channel, leaving the cell in the erased state.

22 Sector erase takes place by tagging the selected sector and
23 raising the associated row lines to a sufficiently high voltage
24 to erase the floating gates to their required erased levels.

25 Because of the low programming currents associated with
26 source side injection (approximately 1-5 microamps/cell),
27 massive parallel programming is made practical, e.g. a full row
28 line of approximately 1000 cells is programmed in a single
29 operation with total current less than approximately 1-5mA,
30 thus providing more than 100 times more efficiency than prior
31 art drain side programming arrays.

32

TABLE 3

State Table & Operating Conditions (Fig. 2a)

Node Operation	Poly 3 (Word line)	Poly 2 (Steering Gate)	Drain	Source
R R STANDBY	0v	0v	Don't care	0v
E E READ SELECTED	5v	0v	2.5v	0v
A L READ UNSELECTED	5v	0v	Don't care	0v
D A				
T				
E				
D				
E R STANDBY	0v	0v	0v	0v
R E				
A L				
S A				
E T ERASE Option 1	12v-22v	0v	0v	0v
E				
D Option 2	5v	-10v to -12v	0v	0v
P R PROGRAM	H1.0v	14-20	0v	5v-8v
R E SELECTED				
O L				
G A				
R T PROGRAM	0v	14-20	0v	5v-8v
A E UNSELECTED	H1.0v	14-20	5v	5v-8v
M D	0v	14-20	5v	5v-8v

Figure 3 is a graph depicting the gate current into poly 1 gate 107 of Fig. 1b (which is not floating in the Figure 3 test device to allow this measurement to be made) as a function of poly 1 gate voltage ($V_{poly\ 1}$) while keeping the select transistor 110 V_{p2} at just above its threshold. In this way most of the potential drop in channel 106 of Figure 1 occurs in channel portion 106-1 underneath gate 109 of select transistor 110, and electrons accelerated in this channel are then injected onto floating gate 107. From Fig. 3 it is seen the hot electron programming injection efficiency of this device is phenomenally high.

Various embodiments of a process suitable for fabricating a structure in accordance with the embodiment of Figures 1a-1d are now described. Reference can also be made to copending U.S. Application Serial No. 323,779 filed March 15, 1989 (now U.S. Patent 5,070,032), and assigned to SunDisk, the assign e

1 of this invention. Reference may also be made to fabrication
2 process steps described earlier in this application. A
3 starting substrate is used, for example a P type substrate (or
4 a P type well region within an N type substrate). A layer of
5 oxide is formed, followed by a layer of silicon nitride. The
6 layer of silicon nitride is then patterned in order to expose
7 those areas in which N+ source and drain regions are to be
8 formed. The N+ source and drain regions are then formed, for
9 example, by ion implantation of arsenic to a concentration of
10 approximately $1 \times 10^{20} \text{ cm}^{-3}$. The wafer is then oxidized in order
11 to form oxide layers 104 and 105 in order to cause source and
12 drain regions 102 and 103 to become "buried". Note that for
13 the embodiment of Figure 2a, this oxidation step is not
14 utilized, as the source and drain regions are not "buried".
15 Rather, the source and drain regions are formed after all
16 polycrystalline silicon layers are formed, in a conventional
17 manner. The remaining portion of the nitride mask is then
18 removed, and the oxide overlying channel regions 106-1 and 106-
19 2 is removed. A new layer of gate oxide overlying channel
20 regions 106-1 and 106-2 is formed, for example to a thickness
21 within the range of 150Å to 300Å and implanted to the desired
22 threshold (e.g. approximately -1v to +1v). Polycrystalline
23 silicon is then formed on the wafer and patterned in order to
24 form floating gate regions 107. If desired, the
25 polycrystalline silicon layer is patterned in horizontal strips
26 (per the orientation of Figure 1a), with its horizontal extent
27 patterned at the same time as the patterning of the second
28 layer of polycrystalline silicon, as will be now described.
29 Following the formation polycrystalline silicon layer 107 at
30 this time, a layer of oxide or oxide/nitride dielectric is
31 formed over the remaining portions of polycrystalline silicon
32 layer 107. A second layer of polycrystalline silicon 108 is
33 then formed and doped to a desired conductivity, for example 30
34 ohms/square. The second layer of polycrystalline silicon is
35 then patterned into vertical strips (again, per the orientation
36 of Figure 1a). If the horizontal extent of polycrystalline
37 silicon layer 107 was not earlier defined, this pattern step is
38 also used to remove the layer of dielectric between the first

1 and second layers of polycrystalline silicon in those areas
2 where the first layer of polycrystalline silicon is to be
3 patterned simultaneously with the patterning of the second
4 layer of polycrystalline silicon. Following the first layer
5 patterning, an additional layer of dielectric is formed on the
6 wafer to form the gate dielectric above channel region 106-1,
7 and above any other areas in the silicon substrate to which the
8 third layer of polycrystalline silicon is to make a gate.
9 These regions can then be implanted to the desired threshold
10 voltage (e.g. approximately 0.5v to 1.5v). The third layer of
11 polycrystalline silicon is for a transistor (ranging from 200Å
12 to 500Å in thickness) then formed and doped to appropriate
13 conductivity, for example 20 ohms/square. Polycrystalline
14 silicon layer 109 is then patterned in order to form word line
15 109.

16 In one embodiment of this invention, polycrystalline
17 silicon layer 107 is patterned to form horizontal stripes and
18 channel stop dopants (e.g. boron) are implanted into the
19 exposed areas therebetween in order to form high threshold
20 channel stop regions between adjacent rows of a memory array.

21 The thickness of the gate dielectric between channel 106-2
22 and polycrystalline silicon floating gate 107 can range from
23 approximately 150 angstroms or less to approximately 300
24 angstroms or more, depending on performance tradeoffs. For
25 increased drive for reading, a thinner gate dielectric is
26 desired while for increased coupling between polycrystalline
27 and silicon control gate 108 and floating gate 107 (helpful
28 during programming) a thicker gate dielectric is desired.

29

30 Second Embodiment

31 Figure 5 is a two-poly embodiment in which programming
32 occurs by taking drain 303 high, for example about 10V while
33 raising control gate 308 just sufficiently so as to turn on
34 select transistor 310. Since this V_{CG} voltage can vary from
35 one device to another it is possible to achieve the optimum
36 injection conditions by keeping V_{CG} at about 3V while raising
37 source (virtual ground) 302 in a sawtooth fashion from about 0

1 to 3 volts and back to 0 again, with a period on the order
2 approximately 1 microsecond.

3 This ensures that at some point along the sawtooth the
4 optimum injection conditions are met. Reference can also be
5 made to European Patent Application Serial No. 89312799.3 filed
6 August 12, 1989. To further enhance programming efficiency, in
7 one embodiment a programming efficiency implant 330 (shown in
8 dotted line) is introduced at the source side. To read the
9 device, its source is 0V, drain is approximately 1.0v and V_{CG}
10 approximately 4.5-5v. To erase we employ poly 1-poly 2
11 tunneling between floating gate 307 in word line 308 at the
12 tunneling zone, consisting of one or more of the floating gate
13 edges, sidewall, corners of the top edge, portions of the top
14 and portions of the bottom, of floating gate 307, associated
15 with a tunnel oxide (400Å-700Å). Erase takes place with V_{CG}
16 approximately 12-22V, $V_D = 0V$, $V_S = 0V$. A capacitive
17 decoupling dielectric (approximately 1500 to 2000Å thick) 340
18 is formed on top of poly 1 to reduce the capacitance between
19 poly 1 and poly 2.

20 In one embodiment of this invention, a high electrical
21 field region is created in the channel far away from the
22 reverse field region located in conventional devices near the
23 drain. This is achieved, for example, by utilizing region 330
24 of increased doping concentration at the boundary between
25 channels 306-1 and 306-2 under floating gate 307. In one
26 embodiment, the width of region 330 is on the order of 0.1
27 microns. A larger dimension for region 330 can be
28 counterproductive, reducing the select transistor drive with no
29 gain in efficiency.

30 Figure 4 depicts the electrical field distribution along
31 channels 306-1 and 306-2 in structures with and without P+
32 doped region 330. In a structure without region 330 and
33 improperly biased select transistor the electron injection can
34 take place in the high field region near drain 303. Because of
35 the vertical field reversal region near drain 303, the
36 resultant injection efficiency is reduced. In a structure with
37 region 330 the injection takes place in the high field region

1 located at region 330, far away from the field reversal region.
2 Because of this, increased injection efficiency is achieved.

3 From the processing side there are three problems which
4 must be addressed properly:

5

6 1. The formation of sufficiently thin/high quality gate
7 dielectric over BN+, which tends to oxidize more quickly than
8 undoped silicon.

9 2. The misalignment between poly 1 and the buried N+ drain
10 diffusion strongly affects the coupling ratios for programming
11 and erase. This can be overcome at the expense of an increase
12 in cell area by not using a virtual ground array, but instead
13 a shared source array.

14 3. This array permits floating gate 307 to completely overlap
15 the buried N+ diffusion in a dedicated source arrangement,
16 eliminating this alignment sensitivity. Unfortunately, this
17 array requires an extra isolation spacing adjacent to the BN+
18 to prevent the poly 1 extension beyond BN+ in the direction
19 away from channel 306-2 to form a transistor in the neighboring
20 cell.

21

22 To achieve small cell size in the buried diffusion
23 direction a channel stop isolation is used between adjacent
24 cells, plus a self-aligned stacked etch to simultaneously
25 delineate poly 2 and poly 1. This is difficult to do without
26 pitting the substrate as well as the exposed BN+ when etching
27 the exposed poly 1 between adjacent cells. This is especially
28 difficult to avoid when etching the decoupling oxide (1500 -
29 2000Å thick on top of poly 1 in order to expose poly 1, since
30 the substrate unprotected by poly 1 also becomes exposed, so
31 that when poly 1 is etched, the substrate in those regions
32 becomes pitted.

33 This will therefore require formation of a thick dielectric
34 region as part of the field isolation process protecting the
35 substrate in the space between the poly 2 word lines. This can
36 be accomplished by using a process as described in U.S. Patent
37 Application Serial No. 323,779, filed March 15, 1989, and
38 assigned to SunDisk, the assignee of this application. This is

1 actually forming trench isolation, but with BN+ abutting this
2 trench, we may experience severe junction leakage as well as
3 loss of a portion of the BN+ conductor. This cell of this
4 second embodiment is attractive because it is double poly, low
5 programming current, very fast programming, programming away
6 from drain junction, small and scalable cell. Cell size is
7 quite attractive as indicated below for three representative
8 geometries:

9 1.0m geometries: cell = $4.0 \times 2.0 = 8.0\text{m}^2$

10 0.8m geometries: cell = $3.2 \times 1.6 = 5.2\text{m}^2$

11 0.6m geometries: cell = $2.3 \times 1.2 = 2.8\text{m}^2$

12 13 Third Embodiment

14 Figure 6 is a cross-sectional view of alternative
15 embodiment of a two poly cell, using source side injection for
16 programming, aided by strong coupling to buried N+ drain 403,
17 which acts also as a second control gate. Erase is by Fowler-
18 Nordheim tunneling to channel 406 through a small thinned oxide
19 region, formed for example to a thickness of about 100\AA , by
20 utilizing a thin polyspacer. These process steps would be as
21 follows: Once the drain oxide is formed (i.e. the oxide above
22 drain 403), a first layer of poly, (approximately 2000\AA to
23 4000\AA thick) is deposited and a thin nitride dielectric is
24 deposited on top. These layers are then etched using a poly 1
25 mask to delineate the lateral extent (as shown in Figure 6) of
26 the poly 1. A second layer of nitride is then deposited and
27 anisotropically etched back to underlying oxide, leaving the
28 initial nitride layer on top of poly 1 plus nitride spacers
29 along the poly 1 sidewalls. This protects the poly 1 sidewall
30 from subsequent oxidation, allowing electrical contact to be
31 made as later described. The exposed oxide layer over the
32 channel portion of the substrate is then stripped and regrown
33 to the 100\AA thickness required for tunneling, while a
34 photoresist masked pattern protects oxide over the exposed, BN+
35 side of the poly 1 from being stripped. The nitride layers
36 surrounding poly 1 prevent oxide from forming on that poly.
37 The thin nitride is then etched off using a highly selective
38 etch which does not attack or degrade the 100\AA tunnel oxide

(e.g. hot phosphoric or plasma etch). This is followed by a second poly deposition which electrically contacts the first poly on its top surface and its sidewalls. This structure is then etched using an anisotropic poly-silicon etch, with etch being terminated with the re-exposure of the oxide layers over substrate beneath the second deposited poly layer. This completes the formation of the poly 1 floating gate stripe shown in Figure 6. The remaining process is similar to that of the second embodiment.

In this embodiment, programming is from hot channel electrons injected from grounded source diffusion 402 with drain 403 held at about +8v and fixed control gate of around 1.5v. Alternatively, programming is performed by hot channel electrons from source diffusion 402 utilizing a sawtooth control gate voltage ranging from 0 volts to a peak voltage approximately 3 volts, as described previously for the second embodiment. Read is achieved with $V_{DS} = 1.5V$, $V_s = 0$, $V_{CG} = +5V$. Erase is achieved with $V_{CG} = -22V$, $V_s = V_d = 0V$. In this embodiment, the poly 2 word line 408 will carry the +5 volts during read and the -22 volts during erase, thereby requiring an X-decoder capable of serving this purpose. Coupling considerations require that $C_{P2P1} > C_{P1D}$, which is unfavorable for programming. Therefore the cell must be optimized for balancing erase against programming by adjusting oxide thicknesses and floating gate threshold to the optimum point. There is less of a problem with pitting the field regions between cells in the poly 1 direction (because poly 1--poly 2 oxide or ONO is thin). This may obviate the need for the additional thick oxide field region described for the second embodiment. However, there is the additional process complexity of forming the thin oxide region and extra space needed to place this thin oxide region sufficiently far from the source diffusion.

Alternative Operating Methods

A number of alternative methods are possible to program the source side injection cells described in the previous embodiments. Strong capacitive coupling (for example, using

thin ONO) is required in the second and third embodiments between poly 2 and drain, and between poly 2 and poly 1, respectively, for programming. During operation, one embodiment applies V_D at 5 to 7v, $V_S = 0$, the control gate voltage V_{CG} is raised to just turn on the control gate channel, and V_{p2} is on the order of about 12 volts or more. Alternatively, the source body effect is used to advantage. In this alternative embodiment, rather than bringing control gate to a specified value to just turn on the channel, the control gate is brought to a value greater than the voltage required to just turn on the channel (e.g., approximately one volt above) and a pull-down circuit is used (e.g., a high impedance resistor or a current sink) for providing approximately 1 μA current flow via source debiasing. Alternatively, the control gate voltage V_{CG} can be operated in a sawtooth fashion from between 0 volts to about +3 volts, as mentioned previously with respect to European patent application serial number 89312799.3.

Multi-bit Cells

In an alternative embodiment of this invention, such as is shown in Figures 9a and 9b, a novel structure is taught including a multi-bit split gate cell, using source side injection programming and using poly-to-poly tunneling for erase. The following describes, in more detail, the operation of one embodiment of such a structure of this invention.

Basic read operation for such a cell consists of applying appropriate control gate bias (e.g. 8v - see TABLE 4) to the unread portion (henceforth for convenience to be termed the transfer portion), while applying the required read control gate bias to the portion being sensed (in multi-state this would be a bias level appropriate to the state being sensed for). In one embodiment, the select gate bias is held at approximately 1.5 volts to keep total cell current limited (e.g. to about 1 microamp), independent of the floating gate conduction level. Alternatively, the select gate bias is maintained at any desired level, e.g. about 5 volts, depending on the current sensing requirements. Similarly, to program a bypass applied on the transfer portion (about 12v) and a

writing potential on the control gate portion (again in multi-state this would be a bias level appropriate to the state being written), with the select gate bias throttled for source side emission (about 1.5v), and the drain bit line (the bit line adjacent the to-be-programmed floating gate) raised to about 5v for programming, with the source bit line (adjacent to transfer portion) grounded.

TABLE 4
OPERATING MODES/CONDITIONS

<u>CONDITION</u>	<u>BL1</u>	<u>CGL2</u>	<u>SG1</u>	<u>CGR2</u>	
<u>BL2</u>					
R STANDBY	X	X	0v	X	X
E READ UNSELECTED	FLOAT	X	1.5v	X	FLOAT
A READ FGL12	0v	READ VREF	1.5v	8v	1.5v
D READ FGR12	1.5v	8v	1.5v	READ VREF	0v
ERASE	0v	0v	VE	0v	0v
P					
R STANDBY	X	X	0v	X	X
O PROG UNSELECTED	FLOAT	X	1.5v	X	FLOAT
G PROG FGL12	5v	PROG VREF	1.5v	12v	0v
R PROG FGR12	0v	12v	1.5v	PROG VREF	5v
A					
M					

NOTES: X - DON'T CARE; VE - OPTIMUM ERASE VOLTAGE (\sim <20v)

Following are some key advantages of the multi-bit cell of this embodiment of this invention:

- (1) Approaches $(2 \times \lambda)^2$ cell size
- (2) Highly self-aligned
- (3) High efficiency source side programming, resulting in lower power and lower voltage requirements, allowing greater parallelism during write
- (4) Attractive for scalability
- (5) Totally immune to overerase

1 This cell can achieve $(2 \times \lambda)^2$ cell size, where λ is the minimum lithographic feature, because (1) each of its lateral component parts, in both its word line and bit line directions, can be formed using this minimum λ feature, and (2) the various critical components are self-aligned to one another, obviating the need to increase cell size to accommodate lithographic overlay registration requirements. For example, viewing along the row or word line direction, the floating gate poly2/1 self-aligned stacks and their underlying channels can be formed using the minimum feature lithographic width (λ), while the transfer channels and bit line diffusions can be simultaneously delineated using the minimum lithographic space between features (also λ), giving it a $(2 \times \lambda)$ minimum pitch capability along this direction. Similarly, looking along the poly2 steering gate in the bit line direction, the channel regions underlying poly1 floating gate and poly3 word line can be formed using the minimum lithographic feature (λ), while the isolation region between word line channels can be formed by the minimum lithographic space (also λ), again achieving the minimum pitch of $(2 \times \lambda)$. In this way, the cell achieves the $(2 \times \lambda)^2$ minimum layout area. It is in fact a self-aligned cross-point cell, the poly2/1 stack and corresponding channel being fully self aligned to the transfer channel and bit line diffusions, and in the orthogonal direction the isolation being self-aligned to the channel areas. When combining this with the low voltage requirement made possible by the source-side injection programming mechanism, this makes it an ideal element for still further scaling (i.e. smaller λ). Finally, its immunity to overerase comes from the following two factors: (1) the presence of the series transistor channel select region, which fully cuts off cell conduction when deselected, independent of degree of erasure, and (2) the source-side injection mechanism itself, which is enhanced with strong overerase, in contrast to the more conventional drain-side programming, which becomes retarded by strong levels of erasure.

1 In one embodiment, rather than the use of 100Å tunneling
2 oxid for the erase operation as in the prior art Ma approach,
3 a thick oxide, geometrically enhanced, poly-to-poly tunneling
4 approach is used, as shown for example in Figures 9a and 9b,
5 where the poly3 word line serves the dual function of cell
6 selection and erase anode (one of the architecture/operational
7 approaches taught in the above-mentioned SunDisk Patent No.
8 5,313,421). Figures 10a and 10b shows the equivalent circuit
9 of this cell/array and TABLE 4 summarizes its operation.

10 The advantages of this embodiment include:

- 11 * Erase unit to follow row line(s), resulting in row oriented
12 sectoring;
- 13 * Avoids need to use negative voltages, erase being
14 accomplished by holding all electrodes at ground, except for
15 the selected sector(s) poly3 word lines, which are raised to
16 erase potential (about or less than 20v);
- 17 * High reliability inherent to thick oxide tunneling
18 implementation; and
- 19 * Improved scalability inherent to the use of the thick
20 interpoly oxide (and consequent reduced parasitic capacitance,
21 both because of the greater thickness and because of the small
22 sidewall vicinity limited tunneling area), combined with the
23 high degree of vertical integration (vertically stacked poly3
24 word line serving the dual role of select gate and erase
25 electrode).

26 Such a cell approach offers the potential for a physically
27 minimal ($4 \times \lambda^2$), highly self aligned, crosspoint cell,
28 which is both very reliable (use of thick oxides and no high
29 voltage junction requirements within memory array), and readily
30 scalable (via the source side injection element and its reduced
31 voltage and more relaxed process control requirements, combined
32 with the inherent scalability of the vertically integrated,
33 thick oxide interpoly erase element). From a physical point of
34 view therefore, a Gigabit (or greater) density level embodiment
35 based on a 0.25μ technology, has a per bit area of
36 approximately $0.25\mu^2$.

37 Despite the series nature of the dual gate cell, a four
38 level multi-stat (two logical bits per floating gate, or four

1 logical bits per dual gate cell) can be implemented. The key
2 requirem nt is that the most heavily programmed state plus bias
3 level of the transfer floating gate's control gate be optimally
4 selected to expose the full multi-state conduction range of the
5 memory floating portion, without introducing read disturb.
6 Based on the above example, a four-level multi-state
7 implementation would give a per bit area approaching $0.1\mu^2$
8 (approximately $0.125\mu^2$).

9 In summary, the above described dual-gate cell based on the
10 thick oxide, row oriented erase approach offers a novel, non-
11 obvious implementation, one that offers significant
12 improvements over the prior art in scalability, reliability and
13 performance.

14 15 Alternative Embodiment Utilizing Negative Steering Cell 16 Operation

17
18 The control gate (or steering) bias voltage level or range
19 of levels for reading constitute a powerful parameter in
20 setting the memory window voltage position and corresponding
21 ranges for the steering element during programming operations
22 and the poly3 control/erase element during erase. By allowing
23 this level or range of levels to go below 0v, this allows
24 shifting up of the floating gate voltage memory window (due to
25 its associated charge) by a proportional amount, governed by
26 the steering gate coupling ratio. The net result is the
27 maximum steering gate voltage level, for both sensing and
28 programming, is reduced by that negatively shifted amount.
29 Similarly, with the steering gate taken below 0v during erase,
30 the maximum erase voltage is also lowered, the amount of which
31 is proportional to the steering gate coupling ratio.

32 An important parameter in determining steering voltage
33 magnitudes is the steering gate coupling ratio, RCG (or R_{21}) =
34 C_{21}/C_{TOT} , where C_{21} is the capacitance between the poly1
35 floating gate and the poly2 steering gate, and C_{TOT} is the
36 total floating gate capacitance. For example, if the net
37 requirement for read plus programming is to capacitively shift
38 the floating gate potential by 10v, then given an RCG of 50%,

1 the steering voltage swing must be scaled up by $1/RCG$, giving
2 a 20v swing. If, on the other hand, RCG is increased to 66.7%,
3 the steering voltage swing drops to 15v, a savings of 5v.
4 Using this 66.7% value, if the read steering bias voltage level
5 (or range) is lowered by 7.5v, the poly3 erase voltage is
6 lowered by $RCG \times 7.5v$, a savings of 5v over the non-lowered bias
7 situation.

8 In order to implement negative steering into an N channel
9 based, grounded substrate memory array, one embodiment utilizes
10 P channel circuitry, capable of going negative of ground, to
11 generate and distribute this bias. In order to support the
12 full steering voltage dynamic range, the N well for such
13 P channel circuitry is biased to the maximum required positive
14 voltage, and the P channel circuitry can thus feed any
15 potential from that value on down to the most negative required
16 (independent of memory array ground). The positive and
17 negative voltage limits are provided from either external
18 supplies or readily generated on chip (for example by N channel
19 based charge pumps for positive bias and P channel for negative
20 bias), since no DC current is required for steering (only
21 capacitive load charging).

22 In one embodiment, a full column oriented array
23 segmentation is implemented to form one sector or a group of
24 row oriented sectors, wherein one sector is read or programmed
25 at any given time. All cells in one sector are erased
26 simultaneously, and one or more sectors can be selected for
27 simultaneous erasure. Column based segmentation breaks a full
28 array into a multiplicity of segmented sub-arrays, thereby
29 eliminating large and/or cumulative parasitics such as
30 capacitance and leakage. Each sub-array has its own set of
31 local bit line diffusions and poly2 steering lines, which are
32 selectively connected by segment select transistor matrixes to
33 corresponding global bit lines and steering lines.

34 Figure 10c exemplifies such a segmentation embodiment,
35 depicting one segment, denoted as SEGMENT I, consisting of N
36 rows of cells (e.g. N equalling 32). For example, each row
37 forms one sector consisting of 2048 dual gate cells or
38 equivalently 4096 floating gate storage elements.

1 Alternatively, a sector can be formed by a group of two or more
2 rows. The long, continuous, global bit lines (typically run in
3 metal) BL_k are selectively connected to the local segment
4 subcolumns through the Segment Bit Line Transfer Select
5 transistors 1001, 1002, driven by the SEG_i lines. Similarly,
6 the long, continuous global steering lines (typically run in
7 metal) S_k are selectively connected to the local segment
8 steering gates through the Steering Drive Transfer Select
9 transistors 2001, 2002, driven by the STD_ODD_i and STD_EVEN_i
10 lines. In this way array segments are isolated from one
11 another, eliminating the large cumulative parasitics of leakage
12 and capacitance, and providing column associated defect and
13 repetitive disturb confinement.

14 Performance can be increased by simultaneously operating
15 on as many cells in one row as possible (where a row may have
16 anywhere from 1K to 4K floating gate memory transistors),
17 thereby maximizing parallelism. Peak power is not a limitation
18 in such implementation, because of the low cell operating
19 currents inherent to this cell approach both during read and
20 programming operations. Consequently, the number of floating
21 gate transistors per row which can be simultaneously operated
22 on is limited only by addressing constraints and segment decode
23 restrictions. For the embodiment shown in Figure 10c, this
24 allows every 4th floating gate to be addressed and operated on,
25 simultaneously, as outlined in TABLE 5, allowing the full row
26 to be addressed and operated on in four passes as follows.

27 During each pass, two adjacent diffusions are driven to
28 drain potential followed by two adjacent diffusions driven to
29 ground, with that bias pattern repeated across the entire row
30 of cells. In this way global drain/source bias is applied in
31 mirrored fashion to every other of the selected cells,
32 resulting in floating gate bias conditions of odd selected
33 cells being reversely applied to those of the even selected
34 cells. Appropriate biases are placed on the global steering
35 lines, as exemplified in TABLE 5, to satisfy the operation of
36 the targeted floating gates as given in TABLE 4, while the
37 local steering lines of the unselected cells are discharged and
38 left isolated from the global steering lines. Once done, the

1 bias conditions for both global bit/ground lines and
2 targeted/untargeted floating gate steering lines are
3 correspondingly interchanged to operate on the other of the
4 floating gate pair within the selected cells. Once this is
5 completed, similar operation is repeated to the alternate set
6 (i.e. previously unselected set) of cells, thereby completing
7 full row programming in four passes.
8

TABLE 5

		GLOBAL BIT LINES							
		BLK-3	BLK-2	BLK-1	BLK	BLK+1	BLK+2	BLK+3	BLK+4
CELLS									
	K-3L K-3R K-2L K-2R								
	K-1R K-1L KR KL								
	K+1L K+1R K+2L K+2R								
	K+3R K+3L K+4R K+4L								
READ	PASS 1	0	1.5	1.5	0	0	1.5	1.5	0
	PASS 2	1.5	0	0	1.5	1.5	0	0	1.5
	PASS 3	0	0	1.5	1.5	0	0	1.5	1.5
	PASS 4	1.5	1.5	0	0	1.5	1.5	0	0
PROGRAM	PASS 1	5	0	0	5	5	0	0	5
	PASS 2	0	5	5	0	0	5	5	0
	PASS 3	5	5	0	0	5	5	0	0
	PASS 4	0	0	5	5	0	0	5	5
ERASE		0	0	0	0	0	0	0	0

37

		SEGMENT _ I LINES				ROW LINES	
		STD EVEN_I	STD ODD_I	SEG_I	SELECTED ROW J LINE	UNSELECTED ROWS	
CELLS K-3L K-3R K-2L K-2R K-1R K-1L KR KL K+1L K+1R K+2L K+2R K+3R K+3L K+4R K+4L							
READ							
PASS 1	SEL UNSEL UNSEL UNSEL	0	10	5	1.5	0	
PASS 2	UNSEL UNSEL UNSEL UNSEL	0	10	5	1.5	0	
PASS 3	UNSEL UNSEL SEL UNSEL	10	0	5	1.5	0	
PASS 4	UNSEL UNSEL UNSEL SEL	10	0	5	1.5	0	
PROGRAM							
PASS 1	SEL UNSEL UNSEL UNSEL	0	14	8	1.5	0	
PASS 2	UNSEL UNSEL UNSEL UNSEL	0	14	8	1.5	0	
PASS 3	UNSEL UNSEL UNSEL UNSEL	14	0	8	1.5	0	
PASS 4	UNSEL UNSEL UNSEL UNSEL	14	0	8	1.5	0	
ERASE	SEL SEL SEL SEL	5	5	5	<20	0	

1 To give an idea of the high speed of this approach with
2 respect to programming, assuming a physical row of 4096
3 floating gate elements, and $10\mu\text{sec}$ per pass for cell
4 programming, this gives an effective programming time of
5 $\sim 10\text{nsec/bit}$ or a raw programming rate of 4096bits per $40\mu\text{sec}$
6 (i.e. per $4 \times 10\mu\text{sec}$) or $\sim 12.5\text{MBytes/sec}$.

7 In order to accommodate the negatively shifted steering in
8 this embodiment, the steering segmentation transistor matrix is
9 implemented in positively biased N well, P channel based
10 circuitry.

11 As indicated above, in order to reduce maximum voltage
12 levels required, it is desirable to keep the steering gate
13 coupling ratio relatively high, for example, greater than
14 approximately 60%, (see Figure 10a for one embodiment of a cell
15 equivalent circuit). In one embodiment, ONO interpoly2/1
16 dielectric (with, for example, an effective tox of 200\AA) is
17 used, combined with a cell structure and process approach
18 (described below), which reduces the parasitic substrate and
19 interpoly3/1 capacitances.

20 Parasitic capacitances to substrate and drain are, in one
21 embodiment, kept small by using a narrow channel structure,
22 bounded by much thicker field oxide regions (such isolation
23 structure is described in U.S. patent 5,343,063). By way of
24 example, a cell with a narrow (for example, about 0.1μ wide),
25 approximately 300\AA thick gate oxide channel region bounded by
26 about 1500\AA thick field regions, whose floating gates are laid
27 out so as to substantially overlap those thick field regions
28 (for example with a total overlap of about 0.3μ), would, in
29 combination with the scaled ONO interpoly2/1, provide steering
30 capacitance magnitudes of around five times larger than those
31 of the floating gate to substrate/drain.

32 In order to reduce the interpoly3/1 capacitance, it must
33 first be noted that in this dual floating gate Flash cell,
34 poly3 crosses two edges of the poly1 floating gate, resulting
35 in approximately double the interpoly3/1 capacitance of cells
36 in which poly3 crosses only a single poly1 edge (for which
37 parasitic coupling ratios are typically around 15%). Although
38 the double edge structure may offer benefits to the erase

1 tunneling element (.g. voltage levels and distributions), its
2 benefit is outweighed by the higher erasing and programming
3 gate voltages needed to offset the associated poorer coupling
4 efficiencies. Therefore, it is desirable to eliminate the
5 capacitive impact of one of these two edges, even if in doing
6 so its erase tunneling contribution is also eliminated. The
7 following discussion describes one embodiment of a process to
8 accomplish this, integrated into the self-aligned diffusion
9 (BN+) formation process.

10 To realize a self-aligned BN+ cell, the BN+ sources/drains
11 must be formed after the poly2/1 stack etch (i.e. self-aligned
12 to poly2) thereby realizing the physically smallest cell. The
13 challenge here is to remove the field oxide locally over the
14 S/D region to allow BN+ As implant, while at the same time
15 preserving sufficiently thick dielectrics surrounding the poly2
16 steering line for poly3 to poly2 high voltage isolation. The
17 following section details the above mentioned exemplary
18 process.

19 In looking at the twin cell in cross-section (see Figures
20 11a-11e for top view and various cross-sections, and in
21 particular Figure 11e), the process strategy to achieve both
22 self-aligned BN+ formation and poly3/1 coupling reduction lies
23 in the ability to separately process the two distinct regions,
24 namely (1) the vertical strip regions associated with the BN+
25 and (2) the vertical strip containing the select channel
26 portions. In so doing, the poly3/1 tunneling edge can be
27 restricted to only form adjacent to the select strip, while
28 completely eliminating its formation along the poly1 edge
29 bordering the BN+ strip.

30 This is accomplished in the following manner (refer to
31 Figure 12a for cross-sections in row line direction, following
32 some of the key process steps. NOTE: the poly3 row lines are
33 defined here to run horizontally, and the BN+ columns to run
34 vertically). By way of example, the following discussion
35 includes representative numbers for dimensions and thicknesses,
36 assuming a 0.25 μ technology (printing minimum lithographic
37 feature size, both width and space, to achieve minimum pitch).

1 Form field oxide 1100 to a thickness of about 1500Å, and
2 etch it into horizontal strips, adding appropriate
3 channel/field implants prior to or at this step. Use an oxide
4 spacer approach to reduce channel width (for example, reduce
5 from about 0.25μ as etched to about 0.1μ post spacer formation,
6 thereby improving control gate coupling). Grow floating gate
7 oxide 1101, (approximately 300Å gate oxide). As shown in
8 Sundisk U.S. patent 5,343,063, the fabrication steps up through
9 the forming of poly1 1102 to a thickness of about 1500Å are
10 performed. Poly1 is then etched into horizontal strips
11 overlying the channel regions plus generous overlap on the
12 field region to either side of the channel. As with channel
13 width, a spacer approach can be used to decrease the etched
14 poly1 spacing, thereby increasing net poly1 overlap of field
15 oxide, or "wings". For example, after the spacer step, poly1
16 spacing is reduced to about 0.1μ, giving poly1 wings of about
17 0.15μ per side - refer to Figure 11b showing a cross-section
18 through the channel along the column direction for an example
19 of poly1 wings over field oxide. Note that because of the
20 narrow channel widths vis a vis the poly1 thickness, poly1 1102
21 will completely fill the trench, resulting in a substantially
22 planar surface. Next form thin ONO 1103 (for example, having
23 about 200Å tox effective) on top of and along edges of the
24 poly1 strips. In an alternative embodiment, a portion of the
25 top film is formed as part of an initially deposited poly1
26 layer stack.

27 Referring to Figures 12a(i) and 12a(ii), deposit a
28 sandwich layer of poly2 1104 (about 1500Å), thick poly3/2
29 isolation oxide 1105 (approximately 2000Å), plus a sufficiently
30 thick etch stopping layer 1106 (to block underlying oxide
31 removal when exposed to oxide type etch), and top oxide layer
32 1107. Using a patterned photoresist masking layer 1108, these
33 are then etched in strips along the column direction, down to
34 the poly1 layer, to form poly2 the steering gate lines. These
35 exposed poly1 regions are overlying the areas to become select
36 channel and BN+.

37 Referring to Figures 12b(i) and 12b(ii), strip previous
38 photoresist and pattern new photoresist layer 1109 to cover and

1 protect exposed poly1 over the select channel regions. Etch
2 exposed poly1 1102 and all its underlying oxide 1101 which
3 cover the to-be-formed BN+ regions. Oxide layer 1107 over etch
4 stopping layer 1106 is used to protect etch stopping layer 1106
5 from being etched by the poly etch as poly1 1102 is being
6 removed. Etch stopping layer 1106 (e.g. thin undoped
7 polysilicon or possibly nitride - must have low etch rate
8 compared to oxide etch rates) is used to prevent that portion
9 of thick poly3/2 isolation oxide 1105 not covered by
10 photoresist 1109 from being etched down as oxide 1101 beneath
11 poly1 1102 is etched away. The oxide etch system used is both
12 highly anisotropic (e.g. RIE) and selective vis a vis the
13 underlying silicon substrate, resulting in negligible etching
14 of that substrate, accommodating the large differences in oxide
15 thicknesses being removed between field oxide (approximately
16 1500Å) and gate oxide regions (approximately 300Å). Following
17 completion of all etching, photoresist 1109 is removed.

18 Referring to Figures 12c(i) and 12c(ii), at this point, an
19 option is to implant and drive a sufficient Boron dose to form
20 a p+ DMOS type doping profile adjacent to the BN+ junction
21 (alternatively, this is the point at which the arsenic BN+ is
22 implanted, but the resulting lateral diffusion makes the
23 floating gate channel unnecessarily short). As shown in Figs.
24 12c(i) and 12c(ii), oxide is formed and reactive ion etched
25 back down to silicon to form sidewall spacers 1110 (about 750Å
26 thick, with the thickness here being determined by interpoly3/2
27 erase high voltage isolation requirements, for example, about
28 25v). The arsenic BN+ strips are then implanted.

29 Referring to Figures 12d(i) and 12d(ii), a new patterned
30 photoresist layer 1111 is added to cover and protect BN+
31 strips. The exposed poly1 over channel strips is etched, to
32 expose the selected channel regions. (Since some of the poly1
33 overlies channel regions and is therefore thicker, while other
34 portions overlie field oxide and is thinner, the same
35 considerations for oxide etch selectivity apply as above, in
36 the oxide over BN+ etching case.)

37 Referring to Figures 12e(i) and 12e(ii), once photoresist
38 1111 is stripped, oxide is formed (e.g. via thermal oxidation

1 or some composite oxide) to simultaneously form the poly1
2 sidewall 1112 and corner interpoly3/1 tunneling oxides 1113
3 (for example, about 350Å), the poly3 gate oxide 1114 over the
4 select channel and oxide 1115 over BN+ (e.g. less than about
5 300Å - the requirement for both of these oxides being they must
6 be sufficiently thick to reliably hold up to the erase voltage
7 to substrate differential). A select transistor threshold
8 adjust implant can be optionally introduced at this time (e.g.
9 increasing channel dopant concentration to raise select V_t , or
10 introducing compensation implant to reduce select V_t).

11 Referring to Figures 12f(i) and 12f(ii), after deposition
12 and patterning of poly3 (which in one embodiment is
13 polysilicide in order to reduce word line delay) the basic dual
14 gate cell structure is complete. In one embodiment of this
15 invention, a high electrical field region is enhanced in the
16 channel far away from the reverse field region located in
17 conventional devices near the drain and source regions. This
18 is achieved, for example, by utilizing regions 1200 of
19 increased doping concentration at the boundary between the
20 channels 1201 and 1202 and transfer channel region 1203. In
21 one embodiment, the width of region 1200 is on the order of 0.1
22 microns.

23 Using the above dimension and film thickness example
24 values, the total floating gate capacitance becomes about 0.4
25 femtoFarads, and coupling ratios are approximately: Steering
26 Gate (R21) 70%; Erase Gate 20%; Floating gate to Substrate &
27 Drain 10%. Although this R21 value may vary from this figure
28 somewhat in that fringing fields from the other terminals are
29 not accounted for, this approximation indicates adequate
30 coupling ratios are achieved in the dual gate cell, even under
31 aggressive cell scaling.

32 A process variant of the above approach, which can reduce
33 further still the erase coupling, is to completely fill the
34 region over BN+ with an oxide, after BN+ formation. This is
35 done, for example, by depositing a sufficiently thick,
36 undensified (and hence easily etched away compared to
37 underlying densified oxide films) oxide layer, patterning
38 photoresist strips over the BN+ to protect it from etching, and

1 etching away the exposed, undensified film over the select
2 channel strips. Following this step and resist removal, the
3 poly3/1 tunnel oxide process proceeds as outlined above, during
4 which time the oxide filler over BN+ is densified.

5 The above approach and its variant outlines one of a
6 number of possible ways to implement the above described dual
7 floating gate cell into the desired array.

8 In summary, several concepts have been introduced to
9 implementing the TWIN FG cell.

10 Fundamental to the cell is its low power source side
11 programming mechanism, and low power row oriented poly-to-poly
12 erase element. Additionally, its independent steering and
13 selection functions, facilitates low power, multi-state read
14 and programming operations.

15 ONO interpoly2/1 is readily integrated to provide a high
16 capacitive coupling, ultra-low leakage steering element. One
17 embodiment uses a full column segment confinement architecture
18 to substantially reduce parasitic bit line capacitance and
19 leakage.

20 A negatively shifted voltage steering implementation
21 allows reduction of maximum voltage ceilings for both the poly2
22 steering lines during programming and the poly3 word/erase
23 lines during erase. Under such implementation, one preferred
24 embodiment for the column segmented array architecture is via
25 an N-well isolated P channel steering selection matrix.

26 High steering ratio is achieved by the narrow channel plus
27 field oxide approach to allow formation of wings. A preferred
28 embodiment is described which reduces the interpoly3/1
29 parasitic as part of a self-aligned BN+ formation process.
30 This replaces the thinner tunneling oxide adjacent one of the
31 two potential tunneling edges with a much thicker isolation
32 oxide. Based on the example used, this approach can give a
33 cell with steering coupling ratio approaching 70%, and
34 parasitic erase coupling down to 20%. Furthermore, based on
35 that example, which uses a 0.25μ technology for the $4\lambda^2$
36 dual floating gate, poly3 word/erase line cell (where λ is
37 the minimum technology feature size), a physical cell area of
38 $0.25\mu^2$ is realizable, which for 8 (16) 1 vel of multi-state

1 translates to an effective cell size approaching $\sim 0.08\mu^2$
2 ($\sim 0.06\mu^2$) per logical bit. These small sizes, around 100 times
3 smaller than physical sizes of cells used in the 4MEG and 8MEG
4 generation of Flash memories, are suitable for building Gigabit
5 density level Flash memories with comparable die sizes and at
6 comparable cost per die.

7 All publications and patent applications mentioned in this
8 specification are herein incorporated by reference to the same
9 extent as if each individual publication or patent application
10 was specifically and individually indicated to be incorporated
11 by reference.

12 The invention now being fully described, it will be
13 apparent to one of ordinary skill in the art that many changes
14 and modifications can be made thereto without departing from
15 the spirit or scope of the appended claims.

1 WHAT IS CLAIMED IS

2

3 1. A memory structure comprising:

4 a source region of a first conductivity type ;

5 a drain region of said first conductivity type;

6 a first channel region of a second conductivity type

7 opposite said first conductivity type, located adjacent said

8 source region;

9 a second channel region of said second conductivity type

10 opposite said first conductivity type, located adjacent said

11 drain region;

12 a transfer channel region of said second conductivity

13 type, located between said first and second channel regions;

14 a first floating gate located above said first channel

15 region;

16 a second floating gate located above said second channel

17 region;

18 a first control gate located above said first floating

19 gate, serving as a steering element associated with said first

20 floating gate;

21 a second control gate located above said second floating

22 gate, serving as a steering element associated with said second

23 floating gate;

24 a third control gate located above said transfer channel

25 region, serving as a control gate of an access transistor, said

26 third control gate also overlying at least a portion of said

27 first and second control gates;

28 a first tunneling zone formed between said first floating

29 gate and said third control gate, and including one or more of

30 edges, side wall, corners of the top edge, portions of the top,

31 and portions of the bottom of said first floating gate; and

32 a second tunneling zone formed between said second

33 floating gate and said third control gate, and including one or

34 more of edges, side wall, corners of the top edge, portions of

35 the top, and portions of the bottom of said second floating

36 gate.

37

38 2. A memory structure comprising:

1 a source region of a first conductivity type;
2 a drain region of said first conductivity type;
3 a first channel region of a second conductivity type
4 opposite said first conductivity type, located adjacent said
5 source region, a portion of said first channel region adjacent
6 said source region being doped to said second conductivity type
7 to a dopant concentration greater than that of said first
8 channel region;
9 a second channel region of said second conductivity type
10 opposite said first conductivity type, located adjacent said
11 drain region, a portion of said second channel region adjacent
12 said drain region being doped to said second conductivity type
13 to a dopant concentration greater than that of said second
14 channel region;
15 a transfer channel region of said second conductivity
16 type, located between said first and second channel regions;
17 a first floating gate located above said first channel
18 region;
19 a second floating gate located above said second channel
20 region;
21 a first control gate located above said first floating
22 gate, serving as a steering element associated with said first
23 floating gate;
24 a second control gate located above said second floating
25 gate, serving as a steering element associated with said second
26 floating gate;
27 a third control gate located above said transfer channel
28 region, serving as a control gate of an access transistor;
29 a first tunneling zone formed between said first floating
30 gate and said third control gate, and including one or more of
31 edges, side wall, corners of the top edge, portions of the top,
32 and portions of the bottom of said first floating gate; and
33 a second tunneling zone formed between said second
34 floating gate and said third control gate, and including one or
35 more of edges, side wall, corners of the top edge, portions of
36 the top, and portions of the bottom of said second floating
37 gate.
38

1 3. A memory structure comprising:
2 a source region of a first conductivity type;
3 a drain region of said first conductivity type;
4 a first channel region of a second conductivity type
5 opposite said first conductivity type, located adjacent said
6 source region;
7 a second channel region of said second conductivity type
8 opposite said first conductivity type, located adjacent said
9 drain region;
10 a transfer channel region of said second conductivity
11 type, located between said first and second channel regions;
12 a first floating gate located above said first channel
13 region;
14 a second floating gate located above said second channel
15 region;
16 a first control gate located above said first floating
17 gate, serving as a steering element associated with said first
18 floating gate;
19 a second control gate located above said second floating
20 gate, serving as a steering element associated with said second
21 floating gate;
22 a third control gate located above said transfer channel
23 region, serving as a control gate of an access transistor;
24 a first tunneling zone formed between said first floating
25 gate and said third control gate, and including one or more of
26 edges, side wall, corners of the top edge, portions of the top,
27 and portions of the bottom of said first floating gate;
28 a second tunneling zone formed between said second
29 floating gate and said third control gate, and including one or
30 more of edges, side wall, corners of the top edge, portions of
31 the top, and portions of the bottom of said second floating
32 gate;
33 a first doped region at the interface of said first
34 channel region and said transfer channel region, said first
35 doped region being doped to said second conductivity type and
36 having a greater dopant concentration than that of said first
37 channel region and said transfer channel region; and

1 a second doped region at the interface of said second
2 channel region and said transfer channel region, said second
3 doped region being doped to said second conductivity type and
4 having a greater dopant concentration than that of said second
5 channel region and said transfer channel region.

6

7 4. A memory array having a plurality of memory cells,
8 comprising:

9 a plurality of diffused lines running in a first
10 direction, serving as source and drain regions of said memory
11 cells, each memory cell having a first channel region located
12 adjacent said source region and a second channel region located
13 adjacent said drain region, and a transfer channel region
14 located between its said first and second channel regions;

15 a plurality of first floating gates, each located above
16 said first channel region of an associated one of said memory
17 cells;

18 a plurality of second floating gates, each located above
19 said second channel region of an associated one of said memory
20 cells;

21 a plurality of first control gate lines, running in said
22 first direction, each located above an associated set of said
23 first floating gates and serving as steering elements
24 associated with each said first floating gate;

25 a plurality of second control gate lines, running in said
26 first direction, each located above an associated set of said
27 second floating gates and serving as steering elements
28 associated with each said second floating gate; and

29 a plurality of row lines, running in a second direction
30 generally perpendicular to said first direction, forming a set
31 of third control gates above said transfer channel regions of
32 each memory cell, overlying at least a portion of associated
33 ones of said first and second control gates and serving as
34 control gates of access transistors of associated memory
35 cells,

36 wherein each of said memory cells is associated with the
37 intersection of one of said diffused lines and one of said row
38 lines, and

1 wherein each memory cell includes a first tunnelling zone
2 formed between said first floating gate and said third control
3 gate, and including one or more of edges, side wall, corners
4 of the top edge, portions of the top, and portions of the
5 bottom of said first floating gate, and

6 wherein each memory cell includes a second tunnelling zone
7 formed between said second floating gate and said third control
8 gate, and including one or more of edges, side wall, corners
9 of the top edge, portions of the top, and portions of the
10 bottom of said second floating gate.

11

12 5. A memory array having a plurality of memory cells,
13 comprising:

14 a plurality of diffused lines running in a first
15 direction, serving as source and drain regions of said memory
16 cells, each memory cell having a first channel region located
17 adjacent said source region, a portion of said first channel
18 region adjacent said source region being doped to said second
19 conductivity type to a dopant concentration greater than that
20 of said first channel region and a second channel region
21 located adjacent said drain region, a portion of said second
22 channel region adjacent said drain region being doped to said
23 second conductivity type to a dopant concentration greater than
24 that of said second channel region, and a transfer channel
25 region located between its said first and second channel
26 regions;

27 a plurality of first floating gates, each located above
28 said first channel region of an associated one of said memory
29 cells;

30 a plurality of second floating gates, each located above
31 said second channel region of an associated one of said memory
32 cells;

33 a plurality of first control gate lines, running in said
34 first direction, each located above an associated set of said
35 first floating gates and serving as steering elements
36 associated with each said first floating gate;

37 a plurality of second control gate lines, running in said
38 first direction, each located above an associated set of said

1 second floating gates and serving as steering elements
2 associated with each said second floating gate ; and
3 a plurality of row lines, running in a second direction
4 generally perpendicular to said first direction, forming a set
5 of third control gates above said transfer channel regions of
6 each memory cell, and serving as control gates of access
7 transistors of associated memory cells,
8 wherein each of said memory cells is associated with the
9 intersection of one of said diffused lines and one of said row
10 lines, and
11 wherein each memory cell includes a first tunnelling zone
12 formed between said first floating gate and said third control
13 gate, and including one or more of edges, side wall, corners
14 of the top edge, portions of the top, and portions of the
15 bottom of said first floating gate, and
16 wherein each memory cell includes a second tunnelling zone
17 formed between said second floating gate and said third control
18 gate, and including one or more of edges, side wall, corners
19 of the top edge, portions of the top, and portions of the
20 bottom of said second floating gate.
21
22 6. A memory array having a plurality of memory cells,
23 comprising:
24 a plurality of diffused lines running in a first
25 direction, serving as source and drain regions of said memory
26 cells, each memory cell having a first channel region located
27 adjacent said source region and a second channel region located
28 adjacent said drain region, and a transfer channel region
29 located between its said first and second channel regions;
30 a first doped region at the interface of each said first
31 channel region and said transfer channel region, said first
32 doped region being doped to said second conductivity type and
33 having a greater dopant concentration than that of said first
34 channel region and said transfer channel region;
35 a second doped region at the interface of each said second
36 channel region and said transfer channel region, said second
37 doped region being doped to said second conductivity type and

1 having a greater dopant concentration than that of said second
2 channel region and said transfer channel region
3 a plurality of first floating gates, each located above
4 said first channel region of an associated one of said memory
5 cells;
6 a plurality of second floating gates, each located above
7 said second channel region of an associated one of said memory
8 cells;
9 a plurality of first control gate lines, running in said
10 first direction, each located above an associated set of said
11 first floating gates and serving as steering elements
12 associated with each said first floating gate;
13 a plurality of second control gate lines, running in said
14 first direction, each located above an associated set of said
15 second floating gates and serving as steering elements
16 associated with each said second floating gate; and
17 a plurality of row lines, running in a second direction
18 generally perpendicular to said first direction, forming a set
19 of third control gates above said transfer channel regions of
20 each memory cell, and serving as control gates of access
21 transistors of associated memory cells,
22 wherein each of said memory cells is associated with the
23 intersection of one of said diffused lines and one of said row
24 lines, and
25 wherein each memory cell includes a first tunnelling zone
26 formed between said first floating gate and said third control
27 gate, and including one or more of edges, side wall, corners
28 of the top edge, portions of the top, and portions of the
29 bottom of said first floating gate, and
30 wherein each memory cell includes a second tunnelling zone
31 formed between said second floating gate and said third control
32 gate, and including one or more of edges, side wall, corners
33 of the top edge, portions of the top, and portions of the
34 bottom of said second floating gate.
35
36 7. A memory structure as in claims 1, 2, 3, 4, 5, or 6
37 wherein said first conductivity type is N and said second
38 conductivity type is P.

1 8. A memory structure as in claim 7 wherein said second
2 conductivity type is provided by boron dopants.

3

4 9. A memory structure as in claims 1, 2, 3, 4, 5, or 6
5 wherein said floating gates comprise a first layer of
6 polycrystalline silicon, said first control gates comprise a
7 second layer of polycrystalline silicon, and said third control
8 gate comprises a third layer of polycrystalline silicon.

9

10 10. A memory structure as in claims 1, 2, 3, 4, 5, or 6
11 which is capable of storing two or more logical states.

12

13 11. A memory array as in claim 10 wherein said floating
14 gates establish one of a plurality of predetermined charge
15 levels for storing a plurality of two or more logical states.

16

17 12. A memory structure as in claims 1, 2, 3, 4, 5, or 6
18 wherein said source region and said drain region comprise
19 buried diffusions.

20

21 13. A memory structure as in claim 12 which further
22 comprises a relatively thick dielectric layer overlying said
23 buried diffusions.

24

25 14. A memory structure as in claims 1, 2, 3, 4, 5, or 6
26 wherein said transfer channel is doped to said second
27 conductivity type to a doped concentration greater than that of
28 first and second channel regions.

29

30 15. A memory structure as in claims 1, 2, 3, 4, 5, or 6
31 wherein said transfer channel is counter doped to said second
32 conductivity type to a net doped concentration less than that
33 of first and second channel regions.

34

35 16. A memory array as in claims 4, 5, or 6 organized into
36 a plurality of sectors, each sector comprising one or more rows
37 and organized such that erasure of all cells of a sector is
38 performed simultaneously.

1 17. A memory array as in claims 4, 5, or 6 organized as
2 a virtual ground array.

3

4 18. A memory array as in claims 4, 5, or 6 wherein said
5 one of first or second floating gates in alternate cells in a
6 given row are verified simultaneously.

7

8 19. A memory array as in claim 18 wherein an entire row
9 is verified utilizing four verification operations.

10

11 20. A memory array as in claims 4, 5, or 6 wherein said
12 one of first or second floating gates of alternate cells in a
13 given row are programmed simultaneously by placing data
14 associated with each memory cell to be programmed on its
15 associated diffused lines.

16

17 21. A memory array as in claim 20 wherein an entire row
18 is programmed utilizing four program operations.

19

20 22. A method for forming a memory structure comprising
21 the steps of:

22 forming a source region of a first conductivity type;

23 forming a drain region of said first conductivity type;

24 forming a first channel region of a second conductivity
25 type opposite said first conductivity type, adjacent said
26 source region;

27 forming a second channel region of said second
28 conductivity type, adjacent said drain region;

29 forming a transfer channel region of said second
30 conductivity type, between said first and second channel
31 regions;

32 forming a first floating gate above said first channel
33 region;

34 forming a second floating gate above said second channel
35 region;

36 forming a first control gate above said first floating
37 gate, serving as a steering element associated with said first
38 floating gate;

1 forming a second control gate above said second floating
2 gate, serving as a steering element associated with said second
3 floating gate;
4 forming a third control gate above said transfer channel
5 region, serving as a control gate of an access transistor, said
6 third control gate also overlying at least a portion of said
7 first and second control gates;
8 forming a first tunneling zone between said first floating
9 gate and said third control gate, and including one or more of
10 edges, side wall, corners of the top edge, portions of the top,
11 and portions of the bottom of said first floating gate; and
12 forming a second tunneling zone between said second
13 floating gate and said third control gate, and including one or
14 more of edges, side wall, corners of the top edge, portions of
15 the top, and portions of the bottom of said second floating
16 gate.

17

18 23. A method for forming a memory structure comprising
19 the steps of:

20 forming a source region of a first conductivity type;
21 forming a drain region of said first conductivity type;
22 forming a first channel region of a second conductivity
23 type opposite said first conductivity type, adjacent said
24 source region, a portion of said first channel region adjacent
25 said source region being doped to said second conductivity type
26 to a dopant concentration greater than that of said first
27 channel region;

28 forming a second channel region of said second
29 conductivity type, adjacent said drain region, a portion of
30 said second channel region adjacent said drain region being
31 doped to said second conductivity type to a dopant
32 concentration greater than that of said second channel region;

33 forming a transfer channel region of said second
34 conductivity type, between said first and second channel
35 regions;

36 forming a first floating gate above said first channel
37 region;

1 forming a second floating gate above said second channel
2 region;
3 forming a first control gate above said first floating
4 gate, serving as a steering element associated with said first
5 floating gat ;
6 forming a second control gate above said second floating
7 gate, serving as a steering element associated with said second
8 floating gate;
9 forming a third control gate above said transfer channel
10 region, serving as a control gate of an access transistor;
11 forming a first tunneling zone between said first floating
12 gate and said third control gate, and including one or more of
13 edges, side wall, corners of the top edge, portions of the top,
14 and portions of the bottom of said first floating gate; and
15 forming a second tunneling zone between said second
16 floating gate and said third control gate, and including one or
17 more of edges, side wall, corners of the top edge, portions of
18 the top, and portions of the bottom of said second floating
19 gate.
20
21 24. A method for forming a memory structure comprising
22 the steps of:
23 forming a source region of a first conductivity type;
24 forming a drain region of said first conductivity type;
25 forming a first channel region of a second conductivity
26 type opposite said first conductivity type, adjacent said
27 source region;
28 forming a second channel region of said second
29 conductivity type, adjacent said drain region;
30 forming a transfer channel region of said second
31 conductivity type, between said first and second channel
32 regions;
33 forming a first floating gate above said first channel
34 region;
35 forming a second floating gate above said second channel
36 region;

1 forming a first control gate above said first floating
2 gate, serving as a steering element associated with said first
3 floating gate;
4 forming a second control gate above said second floating
5 gate, serving as a steering element associated with said second
6 floating gate;
7 forming a third control gate above said transfer channel
8 region, serving as a control gate of an access transistor;
9 forming a first tunneling zone between said first floating
10 gate and said third control gate, and including one or more of
11 edges, side wall, corners of the top edge, portions of the top,
12 and portions of the bottom of said first floating gate;
13 a second tunneling zone between said second floating gate
14 and said third control gate, and including one or more of
15 edges, side wall, corners of the top edge, portions of the top,
16 and portions of the bottom of said second floating gate;
17 forming a first doped region at the interface of said
18 first channel region and said transfer channel region, said
19 first doped region being doped to said second conductivity type
20 and having a greater dopant concentration than that of said
21 first channel region and said transfer channel region; and
22 forming a second doped region at the interface of said
23 second channel region and said transfer channel region, said
24 second doped region being doped to said second conductivity
25 type and having a greater dopant concentration than that of
26 said second channel region and said transfer channel region.
27
28 25. A method for forming a memory array having a
29 plurality of memory cells, comprising the steps of:
30 forming a plurality of diffused lines running in a first
31 direction, serving as source and drain regions of said memory
32 cells, each memory cell having a first channel region located
33 adjacent said source region and a second channel region located
34 adjacent said drain region, and a transfer channel region
35 located between its said first and second channel regions;
36 forming a plurality of first floating gates, each located
37 above said first channel region of an associated one of said
38 memory cells;

1 forming a plurality of second floating gates, each located
2 above said second channel region of an associated one of said
3 memory cells;

4 forming a plurality of first control gate lines, running
5 in said first direction, each located above an associated set
6 of said first floating gates and serving as steering elements
7 associated with each said first floating gate;

8 forming a plurality of second control gate lines, running
9 in said first direction, each located above an associated set
10 of said second floating gates and serving as steering elements
11 associated with each said second floating gate; and

12 forming a plurality of row lines, running in a second
13 direction generally perpendicular to said first direction,
14 forming a set of third control gates above said transfer
15 channel regions of each memory cell, overlying at least a
16 portion of associated ones of said first and second control
17 gates and serving as control gates of access transistors of
18 associated memory cells,

19 wherein each of said memory cells is associated with the
20 intersection of one of said diffused lines and one of said row
21 lines, and

22 wherein each memory cell includes a first tunnelling zone
23 formed between said first floating gate and said third control
24 gate, and including one or more of edges, side wall, corners
25 of the top edge, portions of the top, and portions of the
26 bottom of said first floating gate, and

27 wherein each memory cell includes a second tunnelling zone
28 formed between said second floating gate and said third control
29 gate, and including one or more of edges, side wall, corners
30 of the top edge, portions of the top, and portions of the
31 bottom of said second floating gate.

32

33 26. A method of forming a memory array having a plurality
34 of memory cells, comprising the steps of:

35 forming a plurality of diffused lines running in a first
36 direction, serving as source and drain regions of said memory
37 cells, each memory cell having a first channel region located
38 adjacent said source region, a portion of said first channel

1 region adjacent said source region being doped to said second
2 conductivity type to a dopant concentration greater than that
3 of said first channel region and a second channel region
4 located adjacent said drain region, a portion of said second
5 channel region adjacent said drain region being doped to said
6 second conductivity type to a dopant concentration greater than
7 that of said second channel region, and a transfer channel
8 region located between its said first and second channel
9 regions;

10 forming a plurality of first floating gates, each located
11 above said first channel region of an associated one of said
12 memory cells;

13 forming a plurality of second floating gates, each located
14 above said second channel region of an associated one of said
15 memory cells;

16 forming a plurality of first control gate lines, running
17 in said first direction, each located above an associated set
18 of said first floating gates and serving as steering elements
19 associated with each said first floating gate;

20 forming a plurality of second control gate lines, running
21 in said first direction, each located above an associated set
22 of said second floating gates and serving as steering elements
23 associated with each said second floating gate; and

24 forming a plurality of row lines, running in a second
25 direction generally perpendicular to said first direction,
26 forming a set of third control gates above said transfer
27 channel regions of each memory cell, and serving as control
28 gates of access transistors of associated memory cells,

29 wherein each of said memory cells is associated with the
30 intersection of one of said diffused lines and one of said row
31 lines, and

32 wherein each memory cell includes a first tunnelling zone
33 formed between said first floating gate and said third control
34 gate, and including one or more of edges, side wall, corners
35 of the top edge, portions of the top, and portions of the
36 bottom of said first floating gate, and

37 wherein each memory cell includes a second tunnelling zone
38 formed between said second floating gate and said third control

1 gate, and including one or more of edges, side wall, corners
2 of the top edge, portions of the top, and portions of the
3 bottom of said second floating gate.

4

5 27. A method of forming a memory array having a plurality
6 of memory cells, comprising the steps of:

7 forming a plurality of diffused lines running in a first
8 direction, serving as source and drain regions of said memory
9 cells, each memory cell having a first channel region located
10 adjacent said source region and a second channel region located
11 adjacent said drain region, and a transfer channel region
12 located between its said first and second channel regions;

13 forming a first doped region at the interface of each said
14 first channel region and said transfer channel region, said
15 first doped region being doped to said second conductivity type
16 and having a greater dopant concentration than that of said
17 first channel region and said transfer channel region;

18 forming a second doped region at the interface of each
19 said second channel region and said transfer channel region,
20 said second doped region being doped to said second
21 conductivity type and having a greater dopant concentration
22 than that of said second channel region and said transfer
23 channel region;

24 forming a plurality of first floating gates, each located
25 above said first channel region of an associated one of said
26 memory cells;

27 forming a plurality of second floating gates, each located
28 above said second channel region of an associated one of said
29 memory cells;

30 forming a plurality of first control gate lines, running
31 in said first direction, each located above an associated set
32 of said first floating gates and serving as steering elements
33 associated with each said first floating gate;

34 forming a plurality of second control gate lines, running
35 in said first direction, each located above an associated set
36 of said second floating gates and serving as steering elements
37 associated with each said second floating gate; and

1 forming a plurality of row lines, running in a second
2 direction generally perpendicular to said first direction,
3 forming a set of third control gates above said transfer
4 channel regions of each memory cell, and serving as control
5 gates of access transistors of associated memory cells,

6 wherein each of said memory cells is associated with the
7 intersection of one of said diffused lines and one of said row
8 lines, and

9 wherein each memory cell includes a first tunnelling zone
10 formed between said first floating gate and said third control
11 gate, and including one or more of edges, side wall, corners
12 of the top edge, portions of the top, and portions of the
13 bottom of said first floating gate, and

14 wherein each memory cell includes a second tunnelling zone
15 formed between said second floating gate and said third control
16 gate, and including one or more of edges, side wall, corners
17 of the top edge, portions of the top, and portions of the
18 bottom of said second floating gate.

19

20 28. A method as in claims 22, 23, 24, 25, 26, or 27
21 wherein said first conductivity type is N and said second
22 conductivity type is P.

23

24 29. A method as in claim 28 wherein said second
25 conductivity type is provided by boron dopants.

26

27 30. A method as in claims 22, 23, 24, 25, 26, or 27
28 wherein said floating gates comprise a first layer of
29 polycrystalline silicon, said first control gates comprise a
30 second layer of polycrystalline silicon, and said third control
31 gate comprises a third layer of polycrystalline silicon.

32

33 31. A method as in claims 22, 23, 24, 25, 26, or 27 which
34 is capable of storing two or more logical states.

35

36 32. A method as in claim 31 wherein said floating gates
37 establish one of a plurality of predetermined charge levels for
38 storing a plurality of two or more logical states.

1 33. A method as in claims 22, 23, 24, 25, 26, or 27
2 wherein said source region and said drain region comprise
3 buried diffusions.

4
5 34. A method as in claim 33 which further comprises the
6 step of forming a relatively thick dielectric layer overlying
7 said buried diffusions.

8
9 35. A method as in claims 22, 23, 24, 25, 26, or 27
10 wherein said transfer channel is doped to said second
11 conductivity type to a doped concentration greater than that of
12 first and second channel regions.

13
14 36. A method as in claims 22, 23, 24, 25, 26, or 27
15 wherein said transfer channel is counter doped to said second
16 conductivity type to a net doped concentration less than that
17 of first and second channel regions.

18
19 37. A method as in claims 25, 26, or 27 organized into a
20 plurality of sectors, each sector comprising one or more rows
21 and organized such that erasure of all cells of a sector is
22 performed simultaneously.

23
24 38. A method as in claims 25, 26, or 27 organized as a
25 virtual ground array.

26
27 39. A method as in claims 25, 26, or 27 wherein said one
28 of first or second floating gates in alternate cells in a given
29 row are verified simultaneously.

30
31 40. A method as in claim 39 wherein an entire row is
32 verified utilizing four verification operations.

33
34 41. A method as in claims 25, 26, or 27 wherein said one
35 of first or second floating gates of alternate cells in a given
36 row are programmed simultaneously by placing data associated
37 with each memory cell to be programmed on its associated
38 diffused line.

1 42. A method as in claim 20 wherein an entire row is
2 programmed utilizing four program operations.

3

4 43. A method as in claims 22, 23 or 24 wherein said steps
5 of forming said first floating gate and said first control
6 gate, and said second floating gate and said second control
7 gate comprising the steps of:

8 forming a plurality of polycrystalline silicon strips in
9 a first direction above and insulated from said first and
10 second channel regions;

11 forming a layer of polycrystalline silicon above and
12 insulated from said plurality of polycrystalline silicon
13 strips; and

14 patterning said plurality of polycrystalline silicon
15 strips and said layer of polycrystalline silicon into strips
16 running in a second direction generally perpendicular to said
17 first direction in order to form said first and second floating
18 gates and said first and second control gates.

19

20 44. A method as in claims 25, 26, or 27 wherein said
21 steps of forming said first floating gate and said first
22 control gate, and said second floating gate and said second
23 control gate comprising the steps of:

24 forming a plurality of polycrystalline silicon strips in
25 said second direction above and insulated from said first and
26 second channel regions;

27 forming a layer of polycrystalline silicon above and
28 insulated from said plurality of polycrystalline silicon
29 strips; and

30 patterning said plurality of polycrystalline silicon
31 strips and said layer of polycrystalline silicon into strips
32 running in said first direction in order to form said first and
33 second floating gates and said first and second control gates.

34

35 45. A method as in claim 43 wherein said step of
36 patterning said plurality of polycrystalline silicon strips and
37 said layer of polycrystalline silicon is performed using the

1 minimum feature lithographic width available in the fabrication
2 process.

3

4 46. A method as in claim 44 wherein said step of
5 patterning said plurality of polycrystalline silicon strips and
6 said layer of polycrystalline silicon is performed using the
7 minimum feature lithographic width available in the fabrication
8 process.

9

10 47. A method as in claims 22, 23, 24, 25, 26, or 27
11 wherein said step of forming said transfer channel region and
12 said source and drain regions comprises the step of delineating
13 said transfer channel region and said source and drain regions
14 simultaneously.

15

16 48. A method as in claim 47 wherein said step of
17 simultaneously delineating said transfer channel region and
18 said source and drain regions is performed utilizing the
19 minimum lithographic space between features available in the
20 fabrication process.

21

22 49. A method as in claims 22, 23, 24, 25, 26, or 27 which
23 further comprises the step of forming a tunnel oxide on only
24 the edges of said first floating gate and said second floating
25 gate adjacent to said transfer channel region.

26

27 50. A method as in claims 25, 26, or 27 which further
28 comprises the step of forming a tunnel oxide on only the edges
29 of said first floating gate and said second floating gate
30 adjacent to said transfer channel region to serve as said
31 tunneling zones, comprising the steps of:

32 forming a plurality of polycrystalline silicon strips;

33 forming a second layer of polycrystalline silicon above
34 and insulated from said first layer of polycrystalline silicon;

35 patterning said second layer of polycrystalline silicon to
36 form said plurality of said first and second control gates;

1 patterning said first layer of polycrystalline silicon to
2 remove portions of said first layer of polycrystalline silicon
3 between adjacent pairs of said first and second control gates;
4 forming spacer dielectric on the exposed side walls of
5 said first and second polycrystalline silicon layers;
6 removing exposed portions of said first polycrystalline
7 silicon layer;
8 forming tunnel oxide on the exposed side walls of said
9 first polycrystalline silicon layer; and
10 forming a third layer of polycrystalline silicon.

11

12 51. A method as in claim 50 wherein a portion of said
13 first and second channel regions adjacent said source regions
14 and drain regions, respectively, are doped to concentrations
15 greater than that of the remaining portions of said channel
16 regions prior to said step of forming spacer dielectric and
17 said source and drain regions are formed after said step of
18 forming said spacer dielectric.

19

20 52. A memory array comprising a plurality of segments,
21 each segment including a subarray comprising:
22 a plurality of adjacent bit lines running in a first
23 direction to form a corresponding plurality of columns;
24 a plurality of steering lines running in said first
25 direction;
26 a plurality of word lines running in a second direction
27 generally perpendicular to said first direction to form a
28 corresponding plurality of rows; and
29 a plurality of memory cells, each memory cell being
30 associated with the intersection of one of said bit lines and
31 one of said word lines.

32

33 53. A structure as in claim 52 wherein said word lines
34 serve as said erase lines.

35

36 54. A structure as in claim 53 which includes one or more
37 sectors, each sector containing one or more of said word lines
38 and their corresponding erase lines, each said sector

1 containing a plurality of memory cells capable of being
2 simultaneously erased.

3

4 55. A structure as in claim 53 which includes one or more
5 sectors, each sector containing one or more of said word lines
6 which also serve as erase lines, each said sector containing a
7 plurality of memory cells capable of being simultaneously
8 erased.

9

10 56. A method as in claim 52 which further comprises the
11 step of storing one of two or more logical states in said
12 memory cell.

13

14 57. A memory array as in claim 52 organized as a virtual
15 ground array.

16

17 58. A memory array as in claim 52 which further
18 comprises:

19 a plurality of diffused lines running in said first
20 direction, serving as said bit lines and forming source and
21 drain regions of said memory cells, each memory cell having a
22 first channel region located adjacent said source region and a
23 second channel region located adjacent said drain region, and
24 a transfer channel region located between its said first and
25 second channel regions;

26 a plurality of first floating gates, each located above
27 said first channel region of an associated one of said memory
28 cells;

29 a plurality of second floating gates, each located above
30 said second channel region of an associated one of said memory
31 cells;

32 a plurality of first control gate lines, running in said
33 first direction, each located above an associated set of said
34 first floating gates and serving as those of said steering
35 lines associated with each said first floating gate;

36 a plurality of second control gate lines, running in said
37 first direction, each located above an associated set of said

1 second floating gates and serving as those of said string
2 lines associated with each said second floating gate; and
3 a plurality of row lines serving as said word lines,
4 running in said second direction generally perpendicular to
5 said first direction, forming a set of third control gates
6 above said transfer channel regions of each memory cell,
7 overlying at least a portion of associated ones of said first
8 and second control gates and serving as control gates of
9 access transistors of associated memory cells,

10 wherein each of said memory cells is associated with the
11 intersection of one of said diffused lines and one of said row
12 lines, and

13 wherein each memory cell includes a first tunnelling zone
14 formed between said first floating gate and said third control
15 gate, and including one or more of edges, side wall, corners
16 of the top edge, portions of the top, and portions of the
17 bottom of said first floating gate, and

18 wherein each memory cell includes a second tunnelling zone
19 formed between said second floating gate and said third control
20 gate, and including one or more of edges, side wall, corners
21 of the top edge, portions of the top, and portions of the
22 bottom of said second floating gate.

23

24 59. A memory array as in claim 58 wherein said one of
25 first or second floating gates in alternate cells in a given
26 row are verified simultaneously.

27

28 60. A memory array as in claim 59 wherein an entire row
29 is verified utilizing four verification operations.

30

31 61. A memory array as in claim 58 wherein alternate cells
32 in a given row are programmed simultaneously by placing data
33 associated with each memory cell to be programmed on its
34 associated bit line.

35

36 62. A memory array as in claim 61 wherein an entire row
37 is programmed utilizing four program operations.

38

- 1 63. A structure as in claims 1, 2, 3, 4, 5, 6, or 52
2 which further comprises steering bias circuitry capable of
3 providing steering bias voltage levels less than zero.
4

1/22

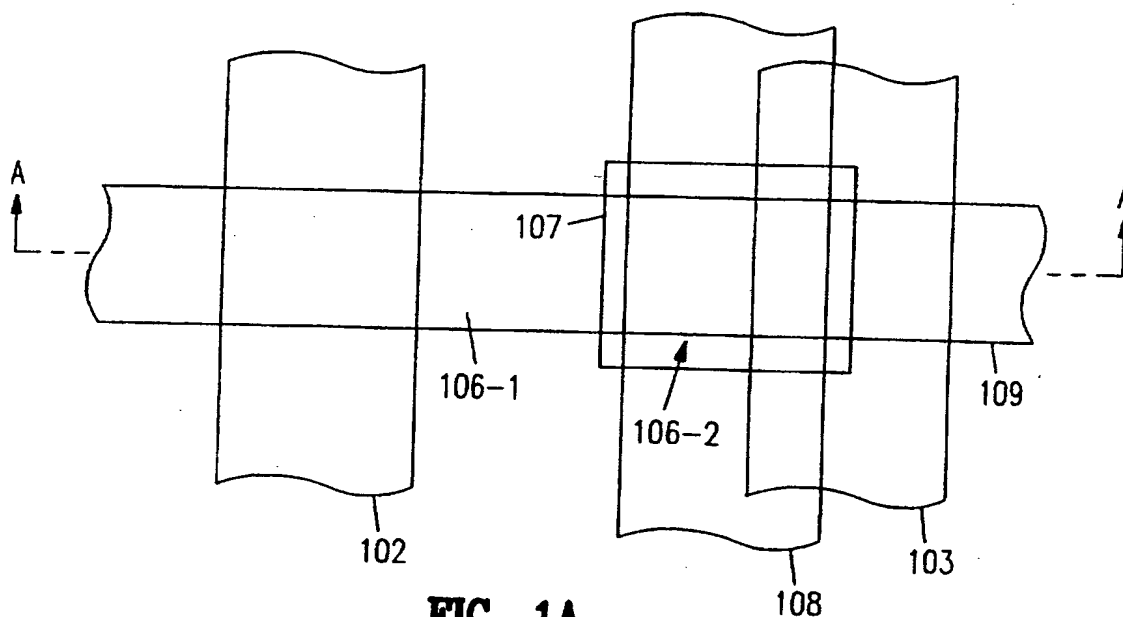


FIG. 1A

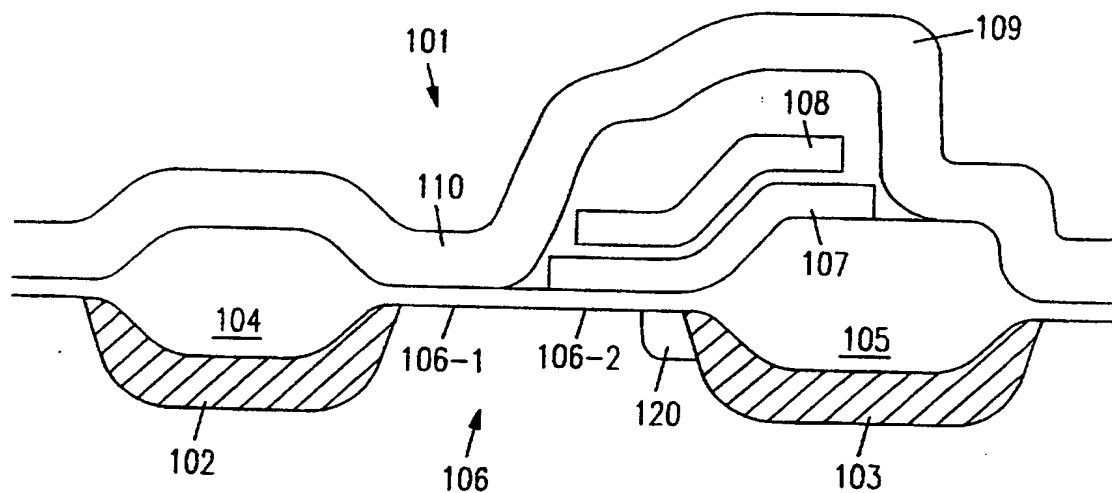


FIG. 1B

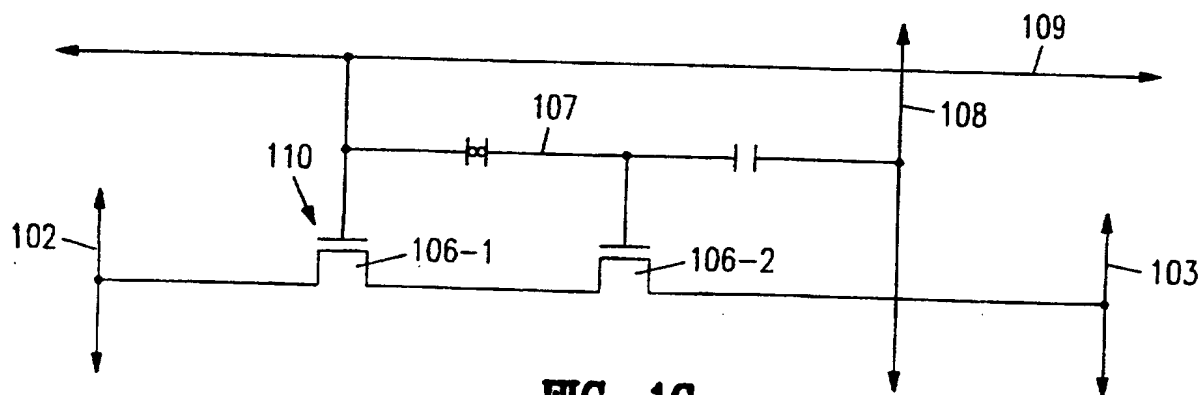
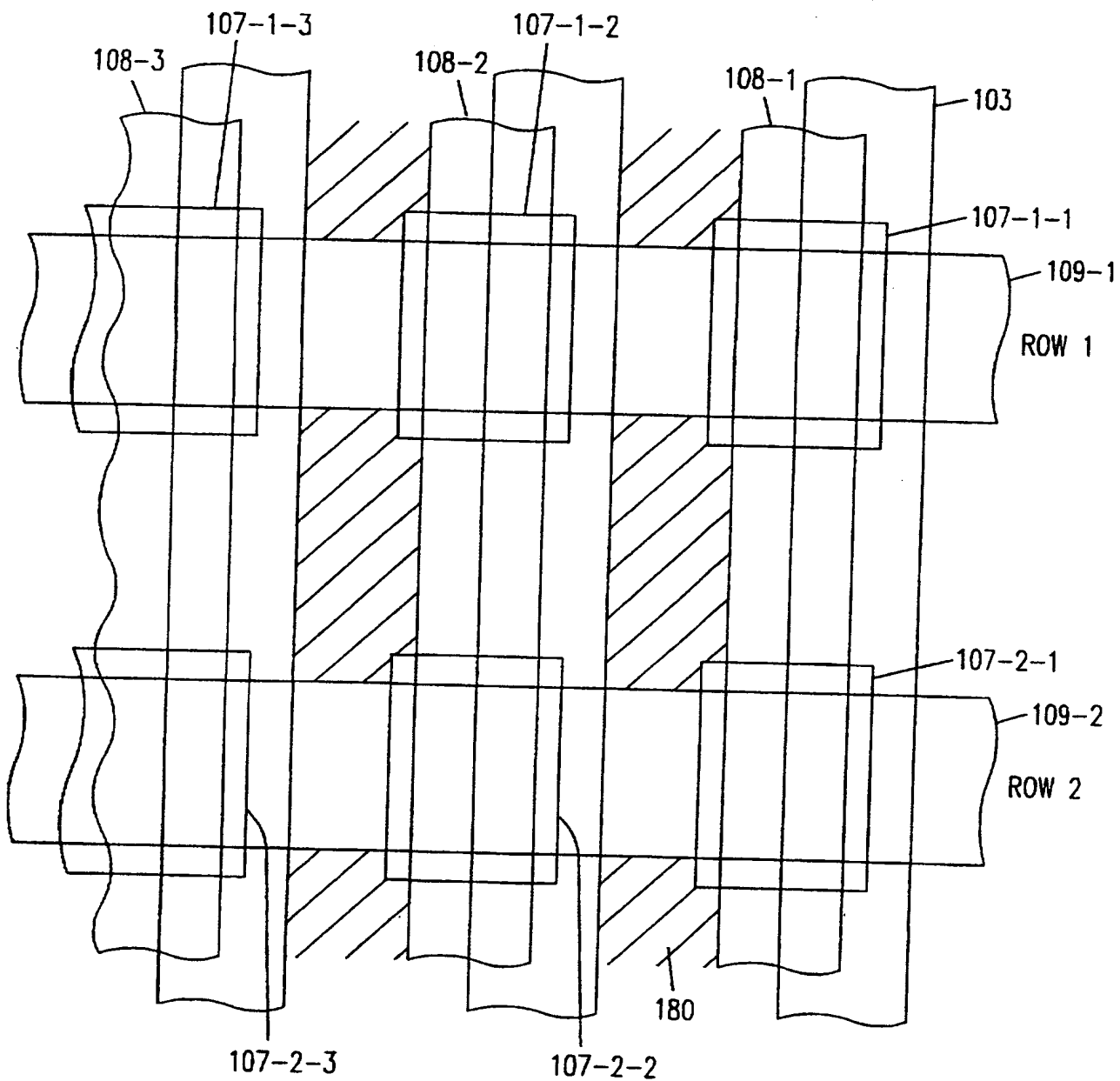


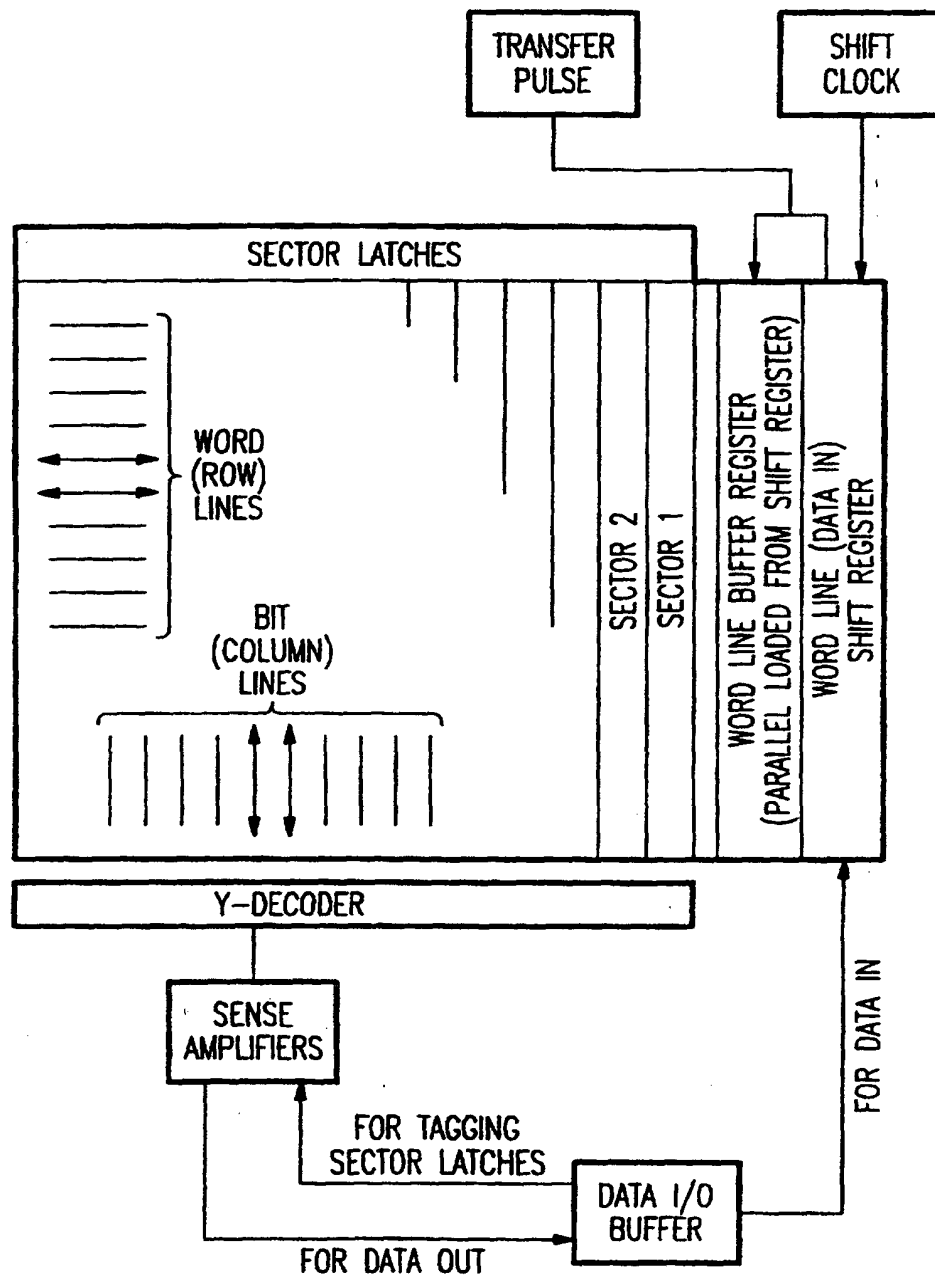
FIG. 1C
SUBSTITUTE SHEET (RULE 26)

2/22

**FIG. 1D**

SUBSTITUTE SHEET (RULE 26)

3/22

**FIG. 1E**

SUBSTITUTE SHEET (RULE 26)

4/22

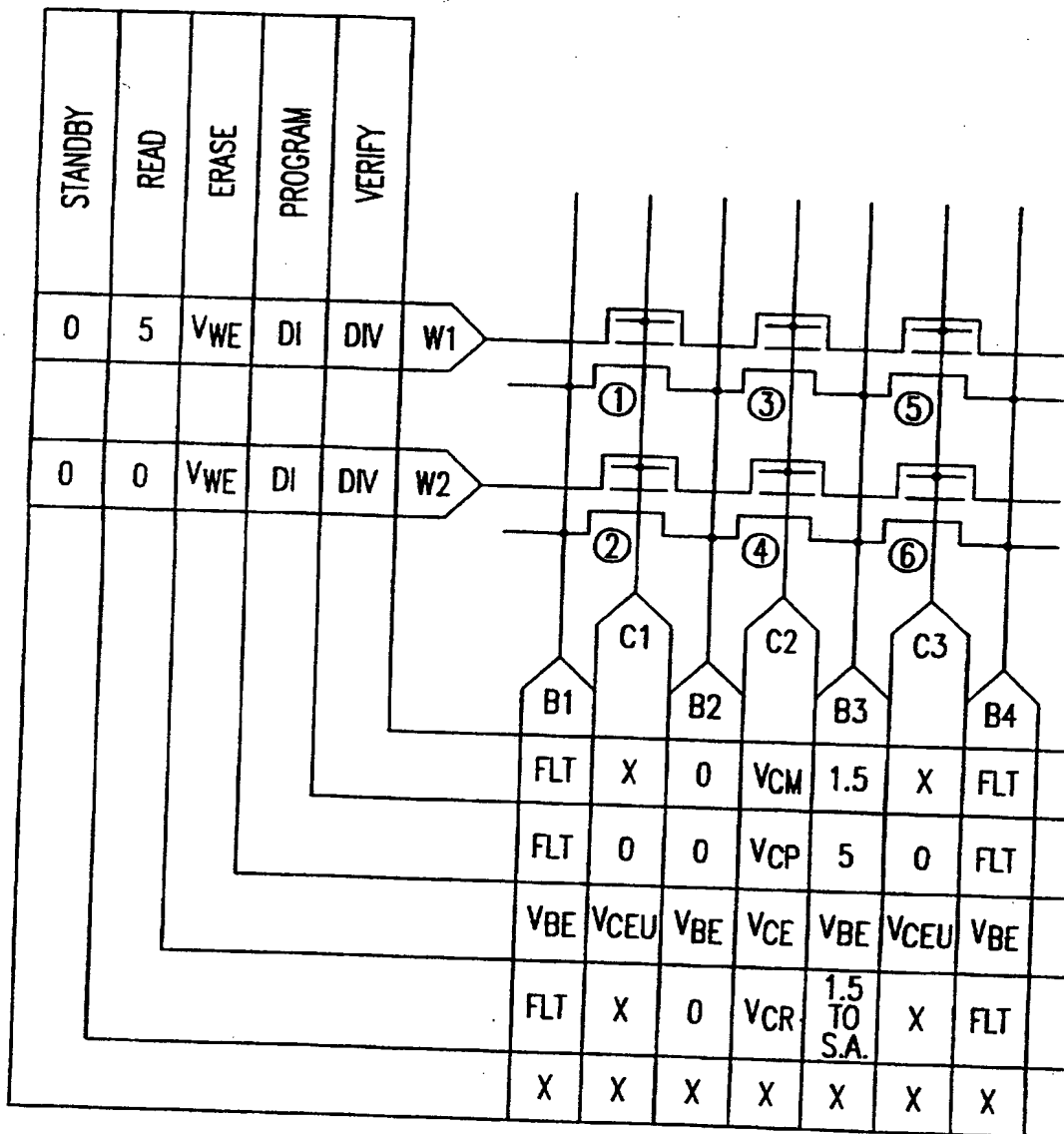


FIG. 1F

SUBSTITUTE SHEET (RULE 26)

5/22

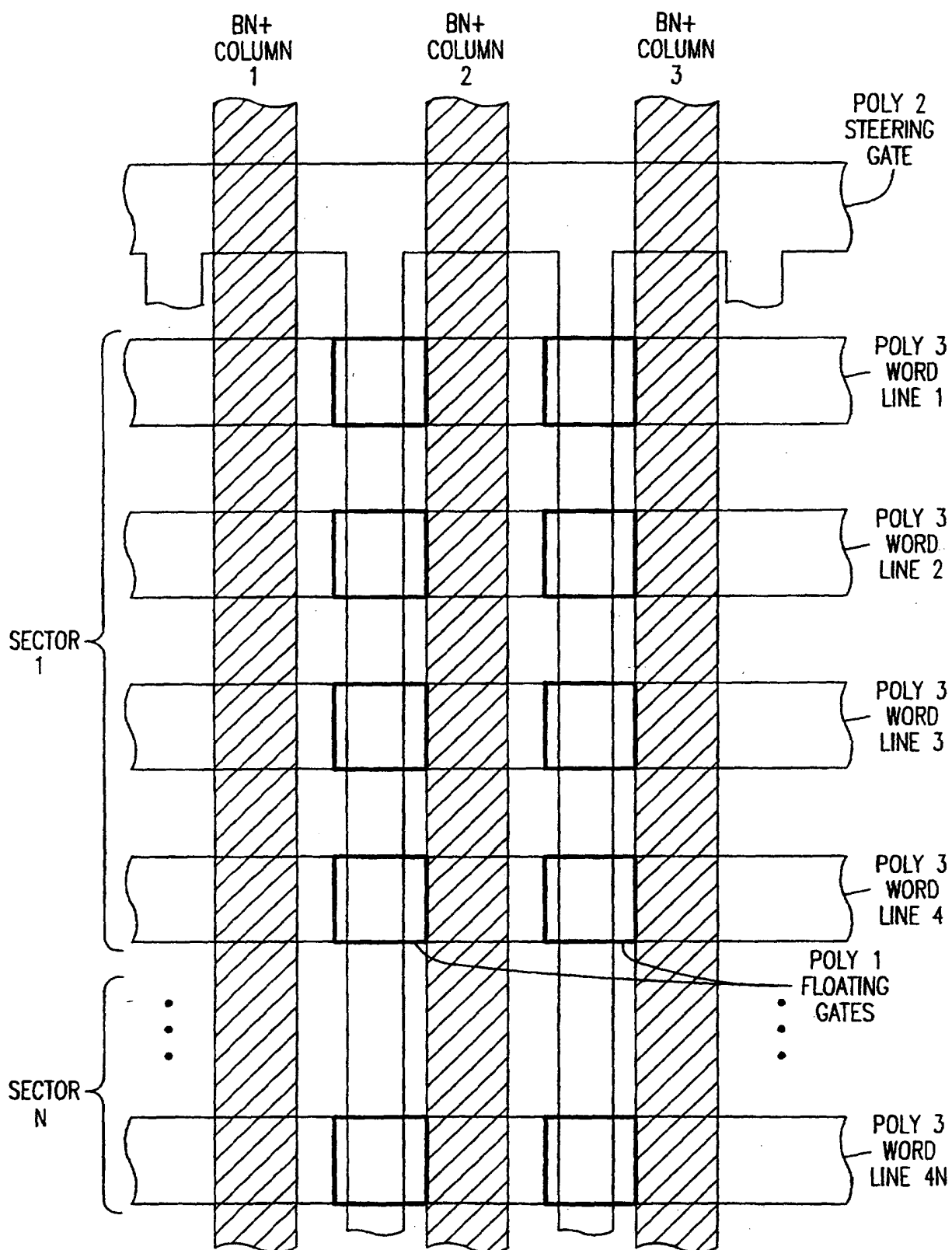


FIG. 1G
SUBSTITUTE SHEET (RULE 26)

6/22

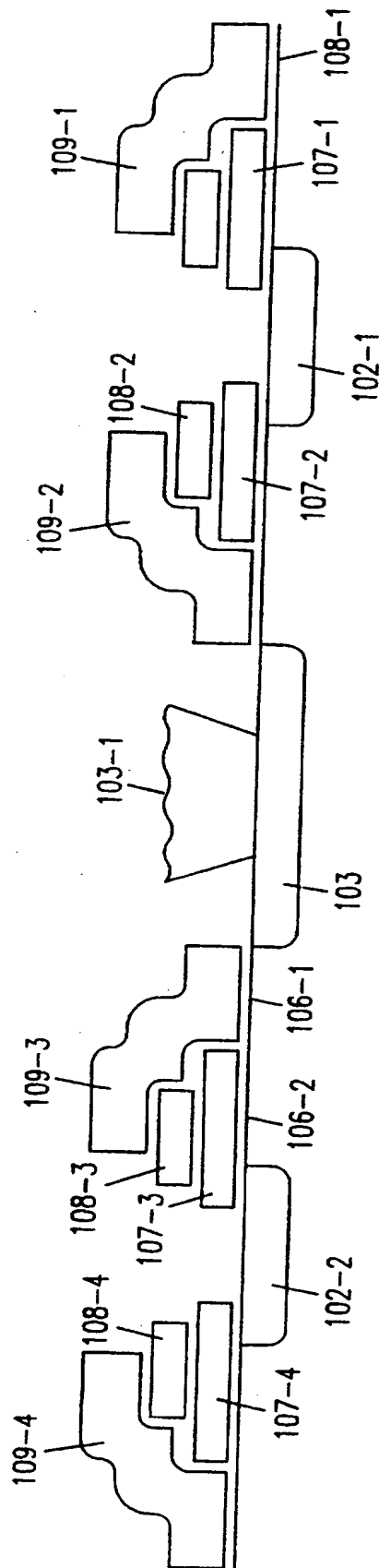
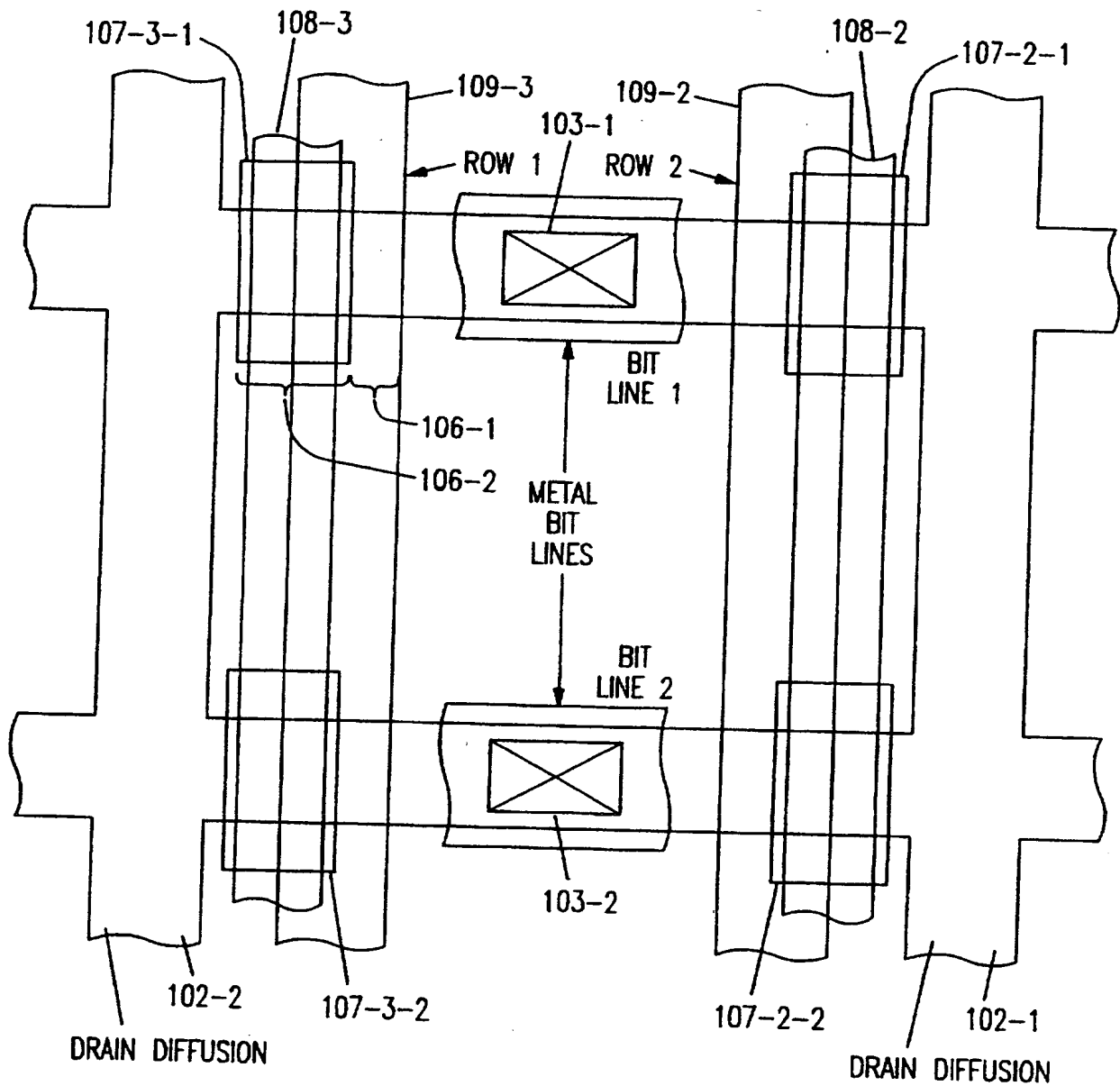


FIG. 2A

7/22

**FIG. 2B**

SUBSTITUTE SHEET (RULE 26)

8/22

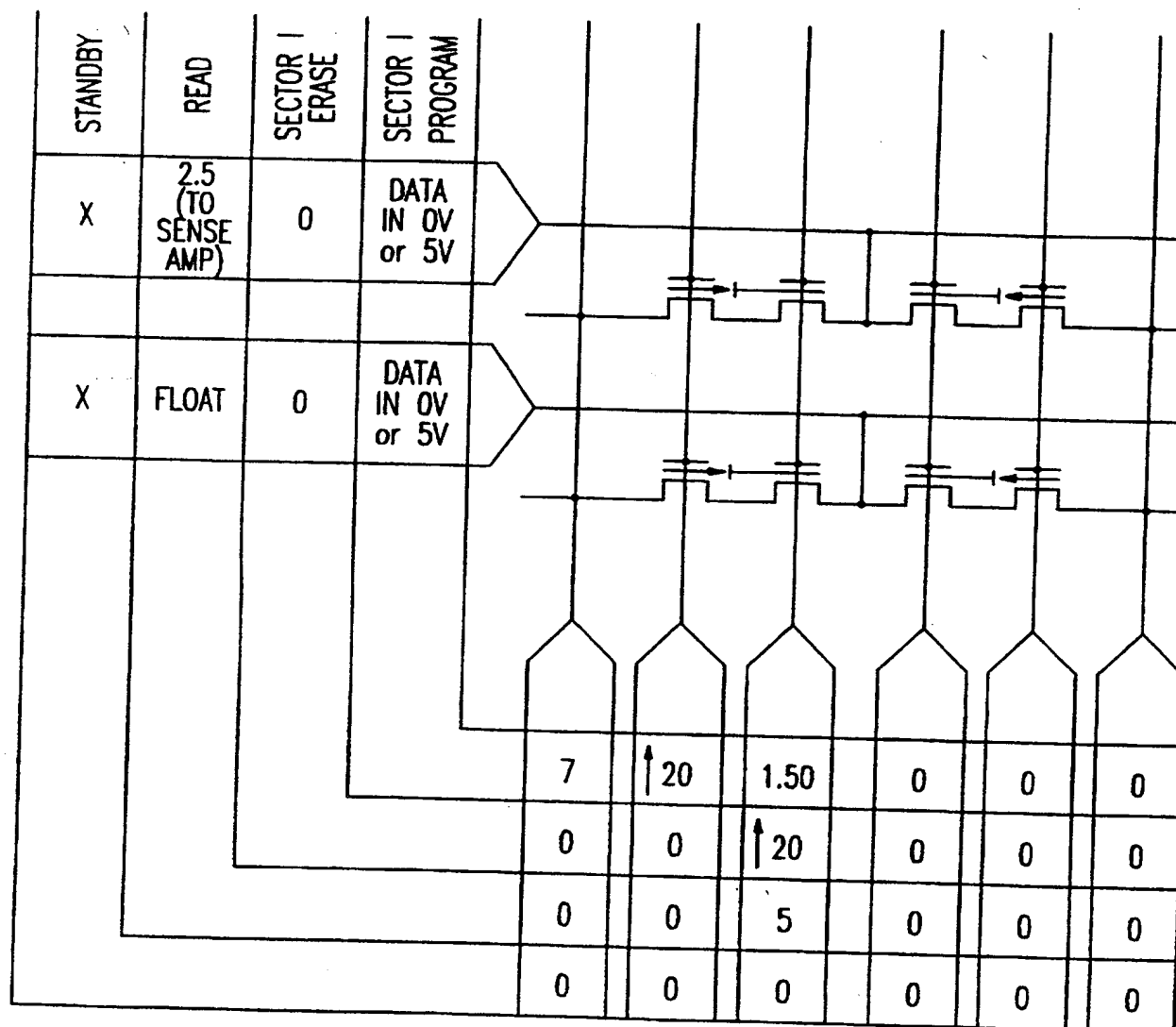


FIG. 2C

9/22

VARIABLE:

VP1 -CH4

LINEAR SWEEP

START .0000V

STOP 18.000V

STEP .3000V

CONSTANTS:

VS -CH1 .0000V

VP2 -CH2 2.4000V

VD -CH3 6.0000V

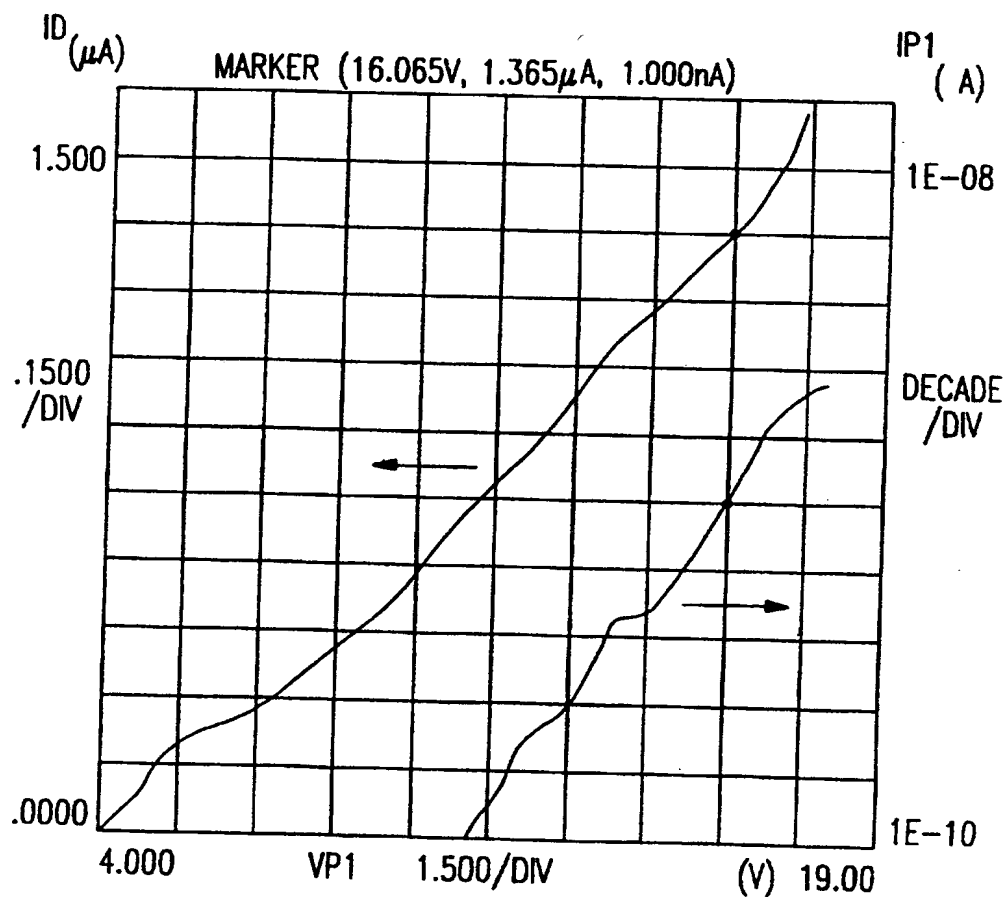
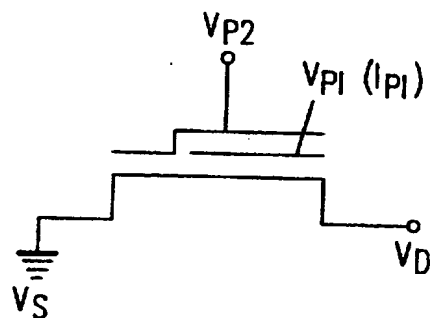


FIG. 3

SUBSTITUTE SHEET (RULE 26)

10/22

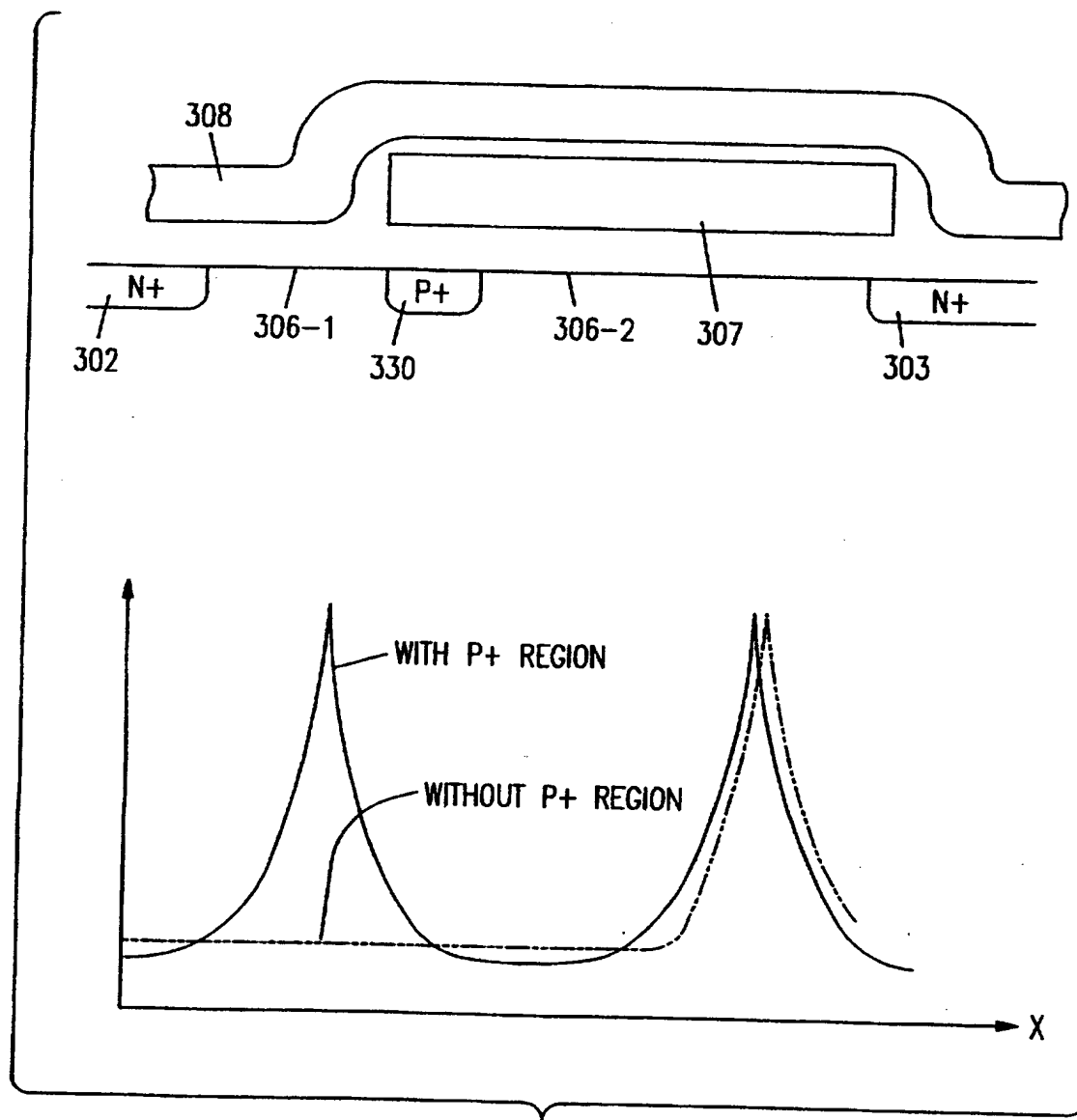


FIG. 4

SUBSTITUTE SHEET (RULE 26)

11/22

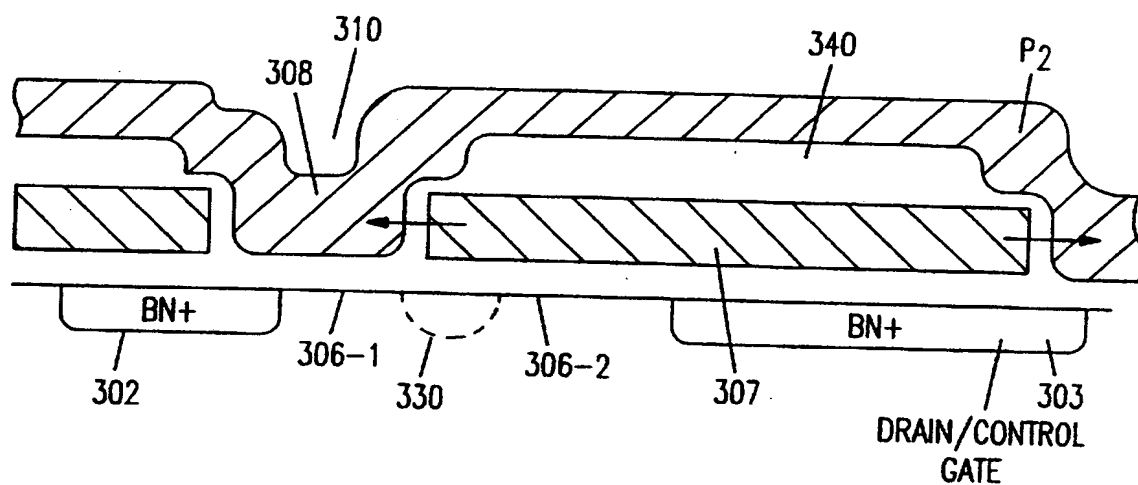


FIG. 5

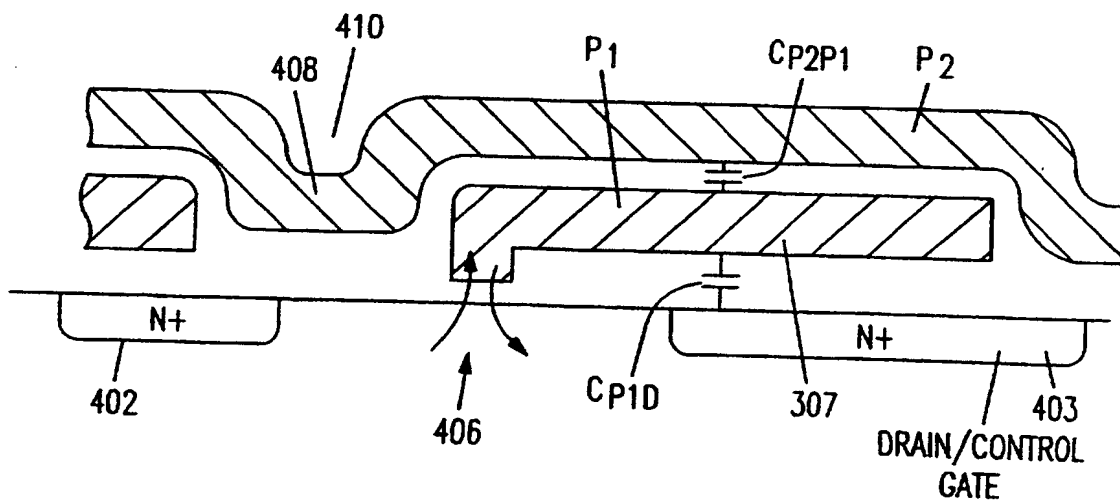


FIG. 6

SUBSTITUTE SHEET (RULE 26)

12/22

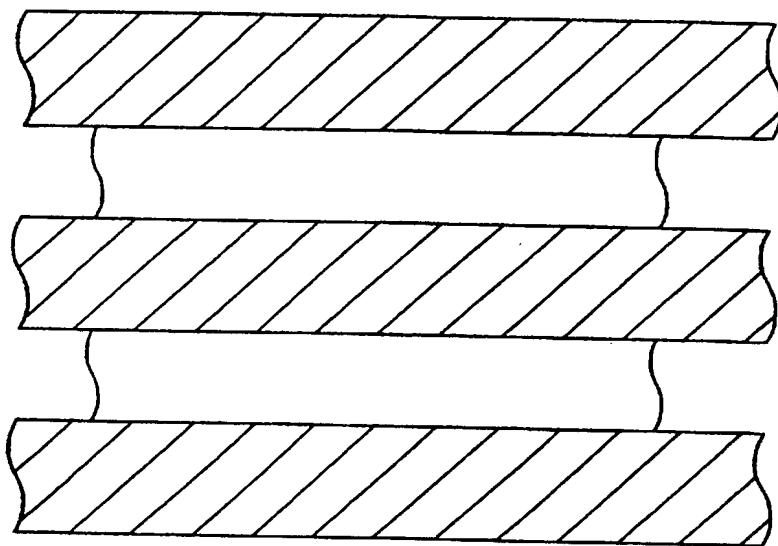


FIG. 7A

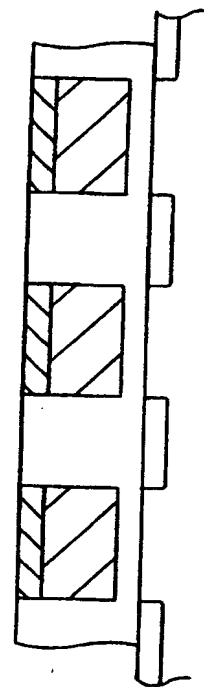


FIG. 7B

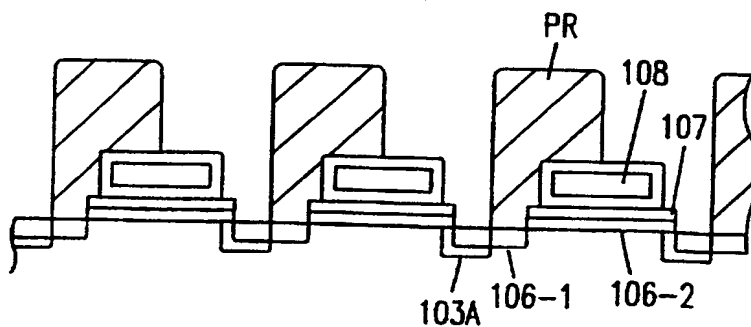


FIG. 8

SUBSTITUTE SHEET (RULE 26)

13/22

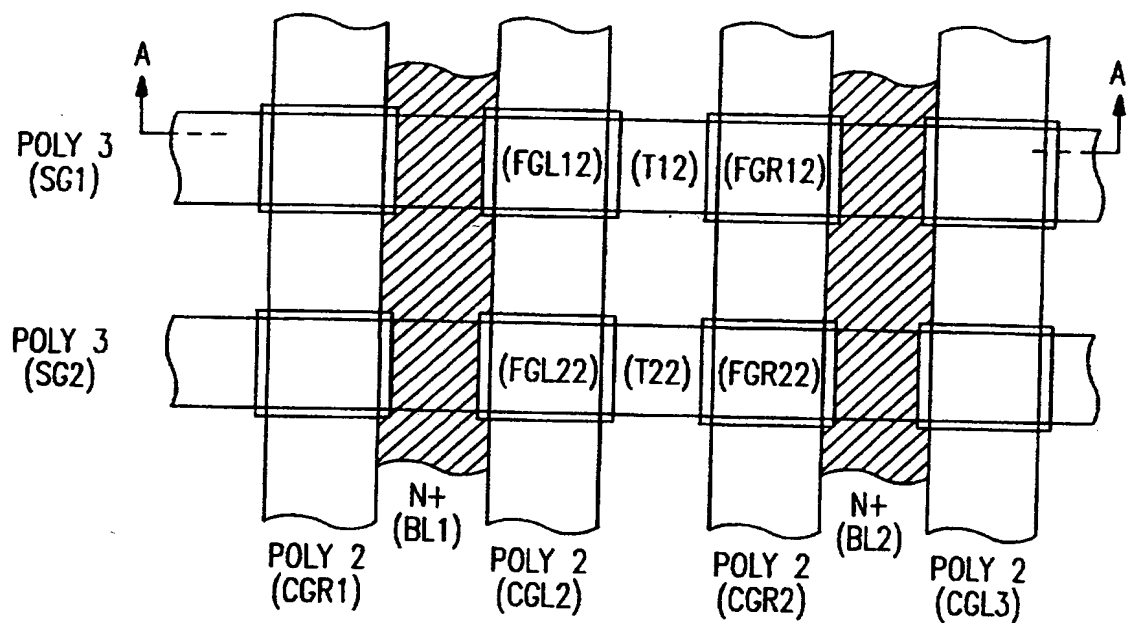


FIG. 9A

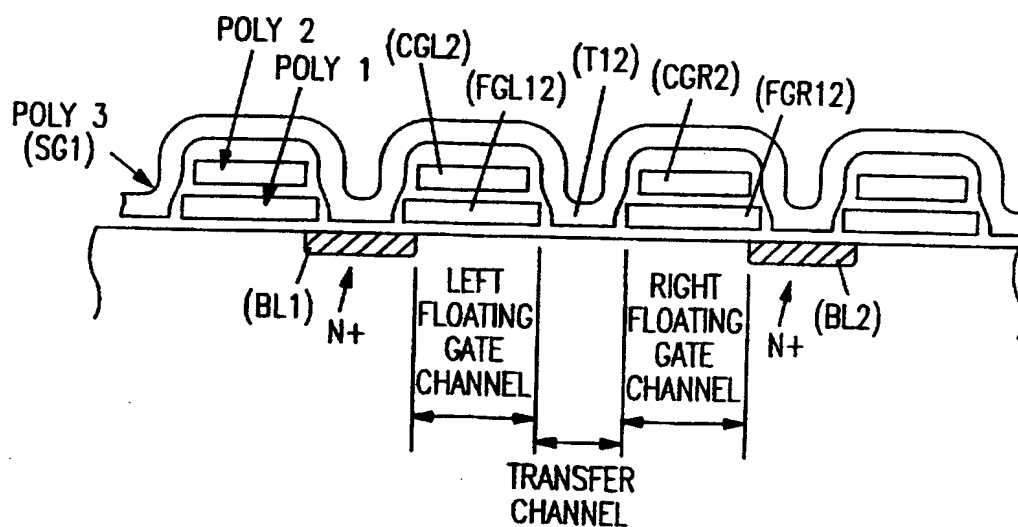


FIG. 9B

SUBSTITUTE SHEET (RULE 26)

14/22

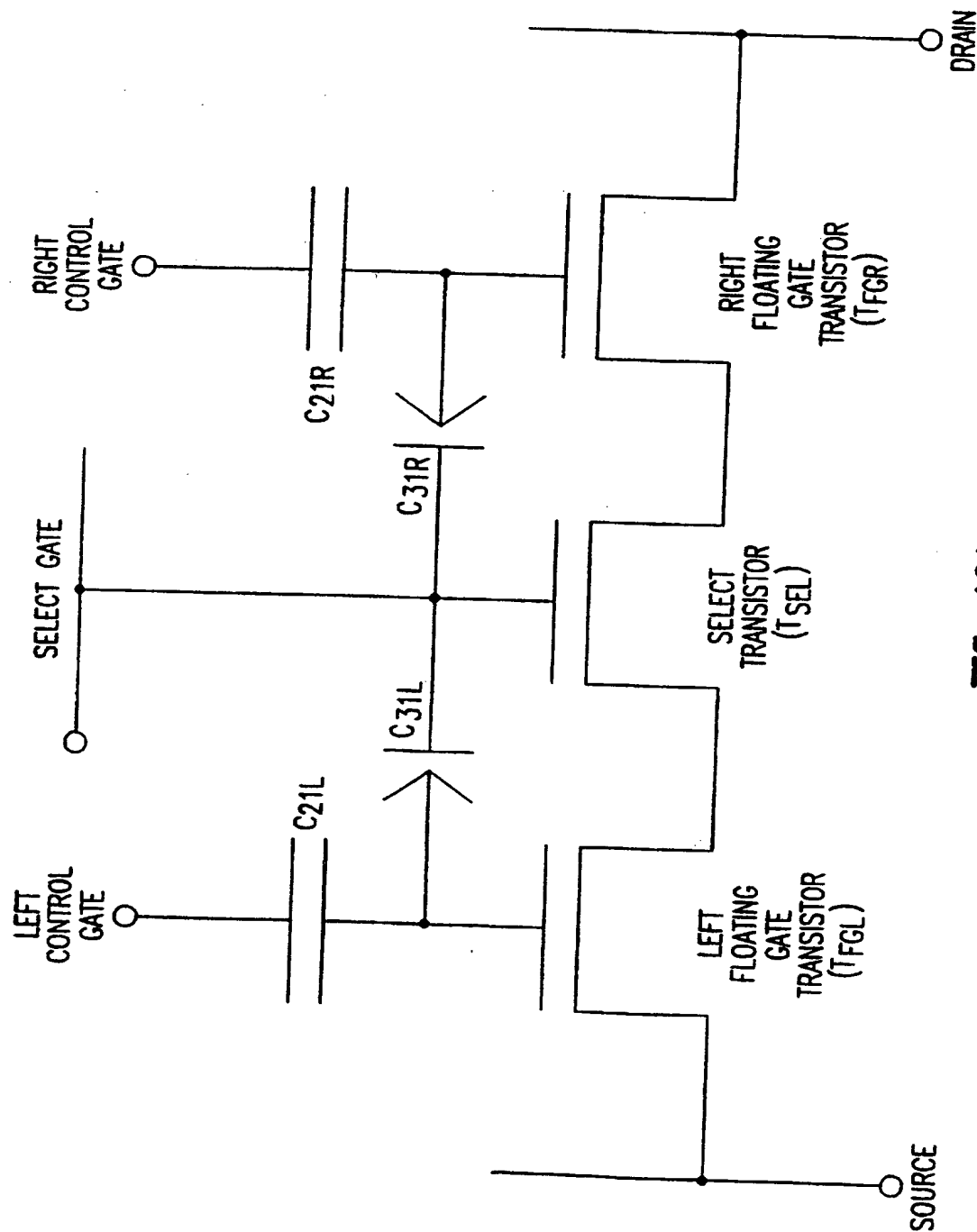


FIG. 10A

SUBSTITUTE SHEET (RULE 26)

15/22

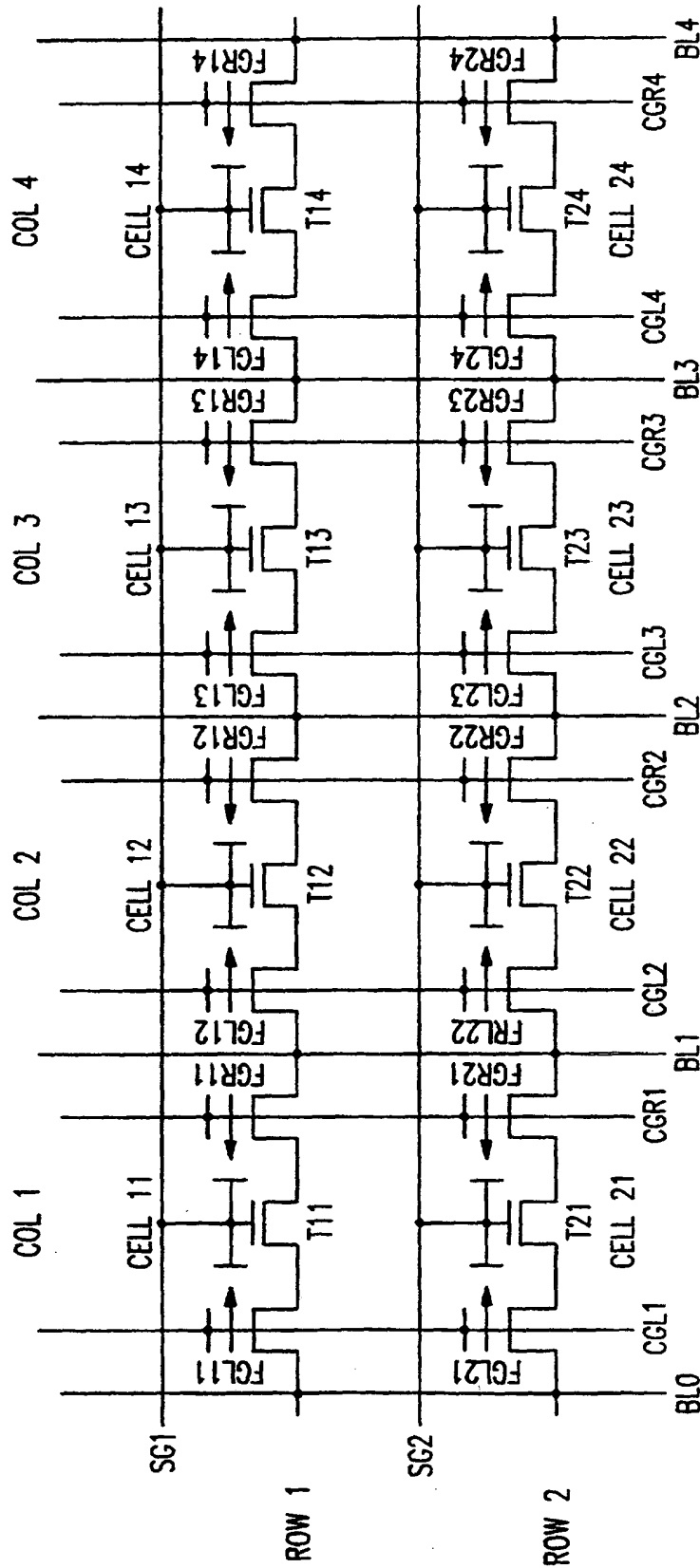


FIG. 10B

SUBSTITUTE SHEET (RULE 26)

16/22

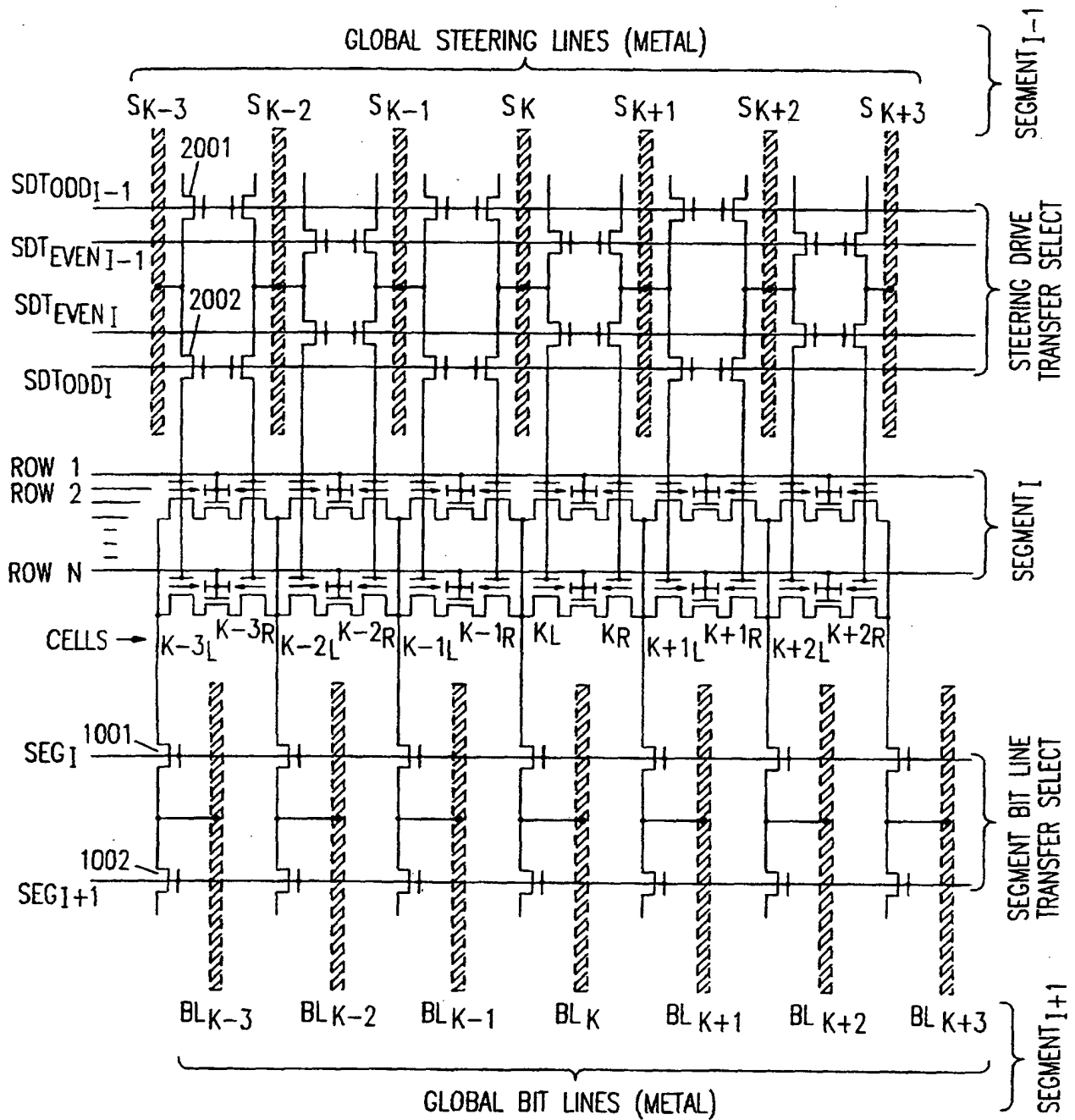


FIG. 10C
 SUBSTITUTE SHEET (RULE 26)

17/22

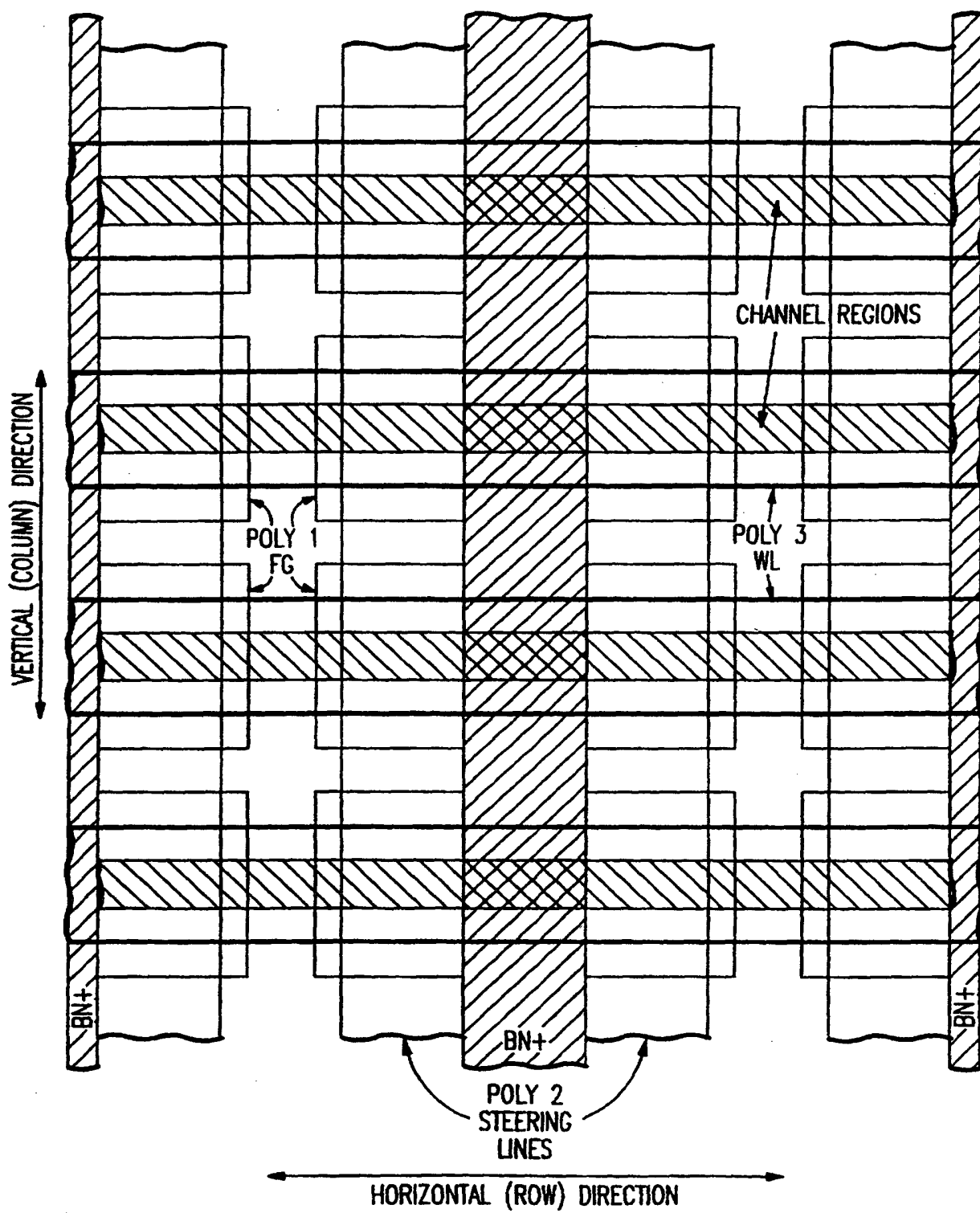


FIG. 11A
SUBSTITUTE SHEET (RULE 26)

18/22

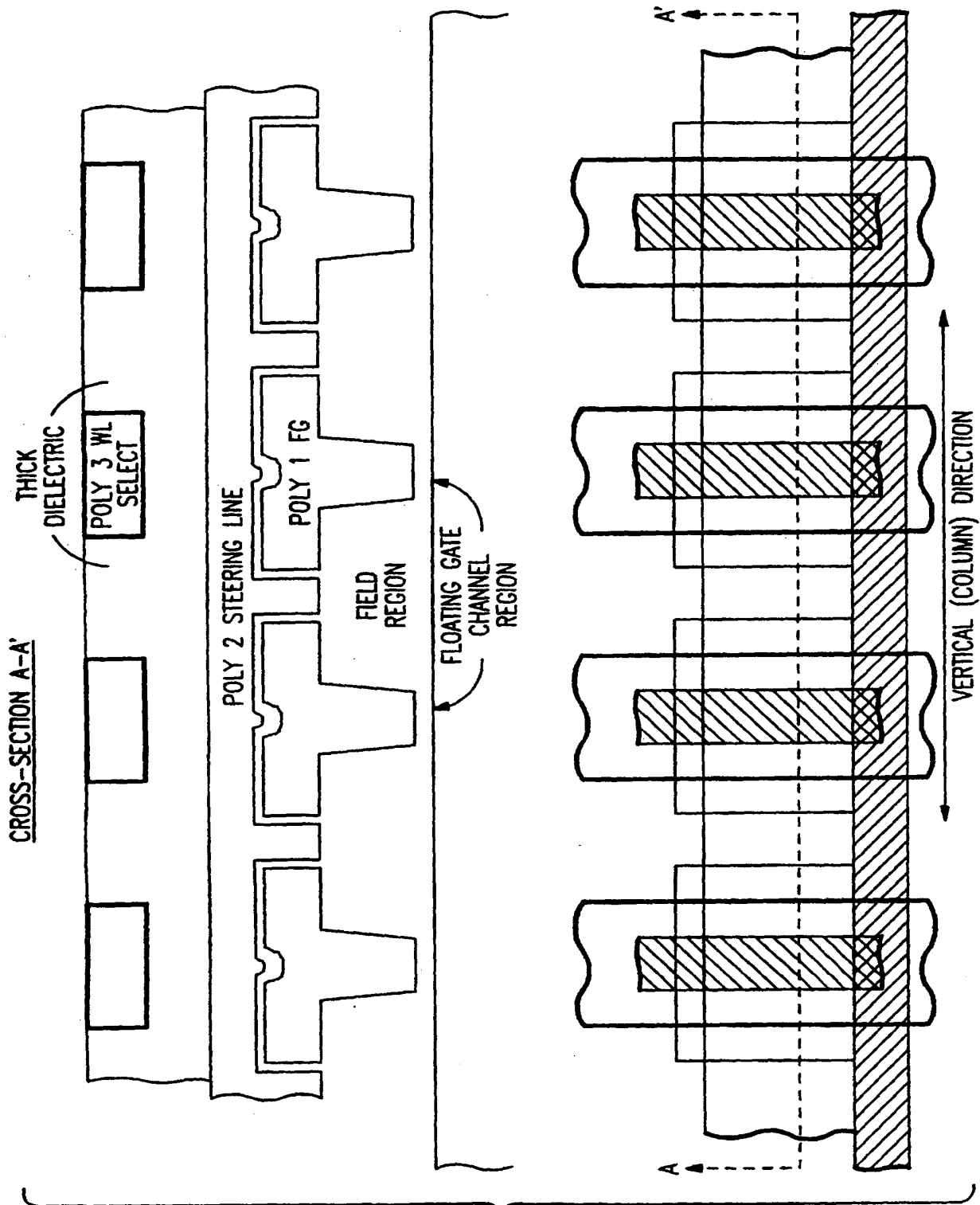


FIG. 11B

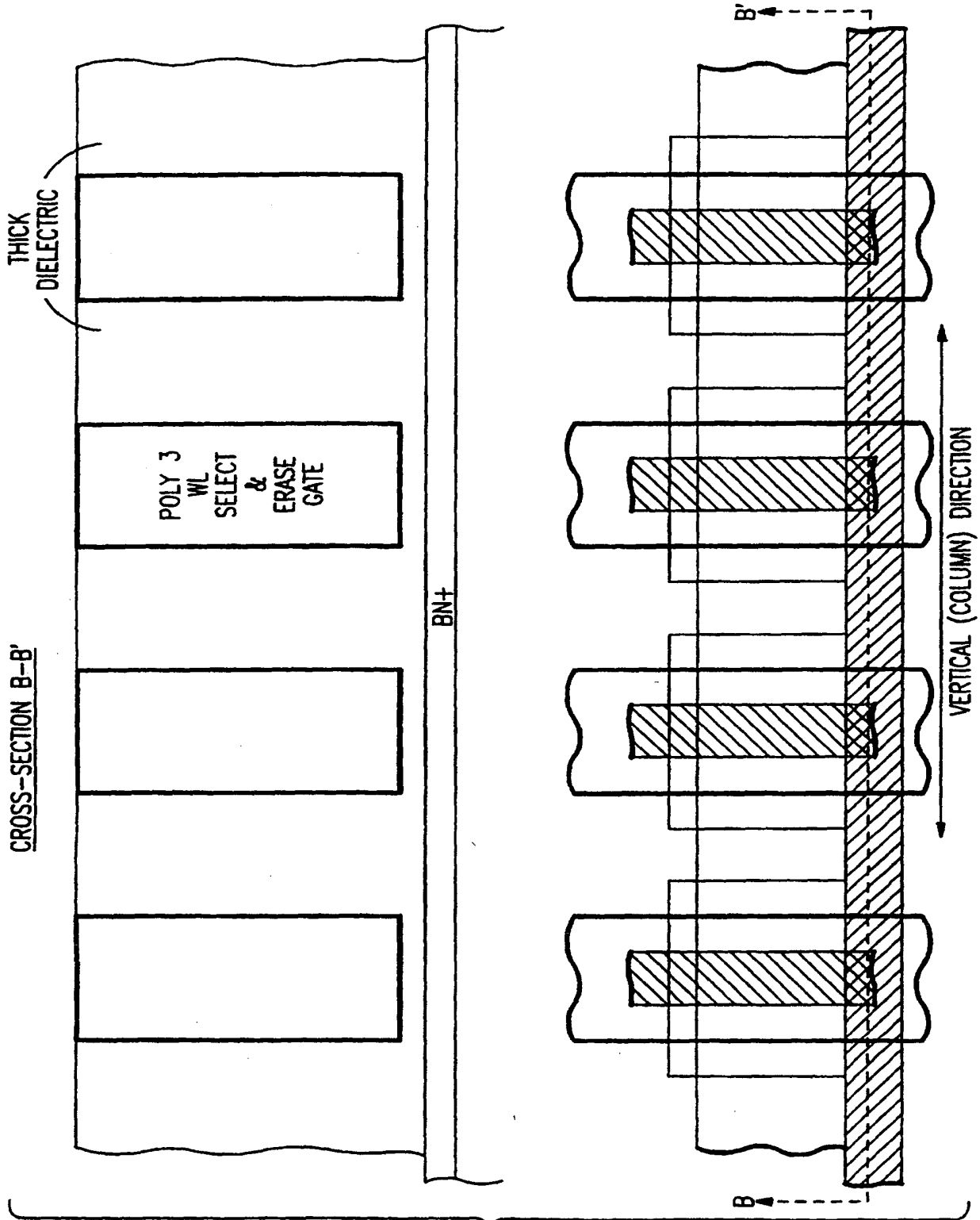


FIG. 11C

SUBSTITUTE SHEET (RULE 26)

20/22

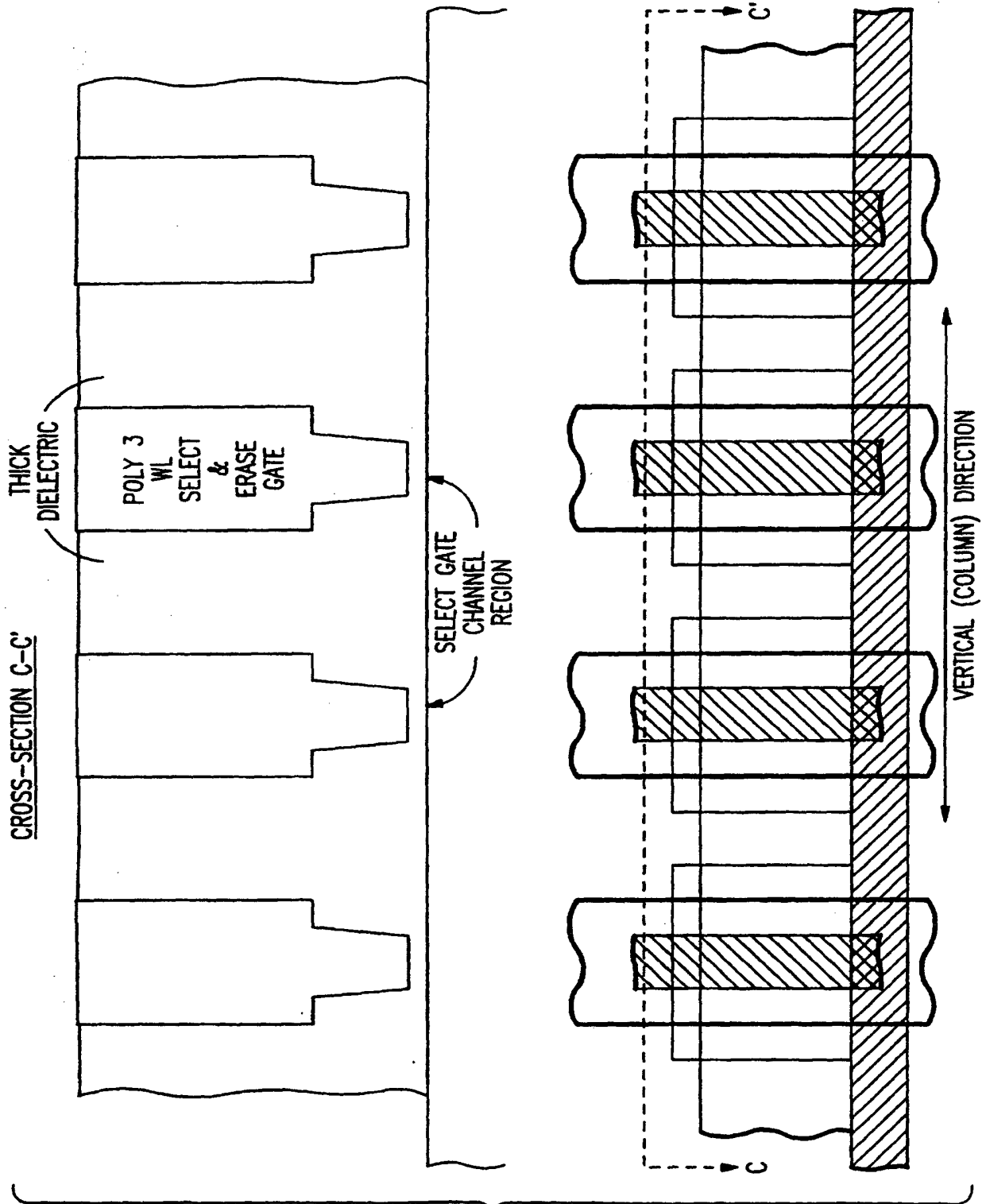


FIG. 11D

SUBSTITUTE SHEET (RULE 26)

21/22

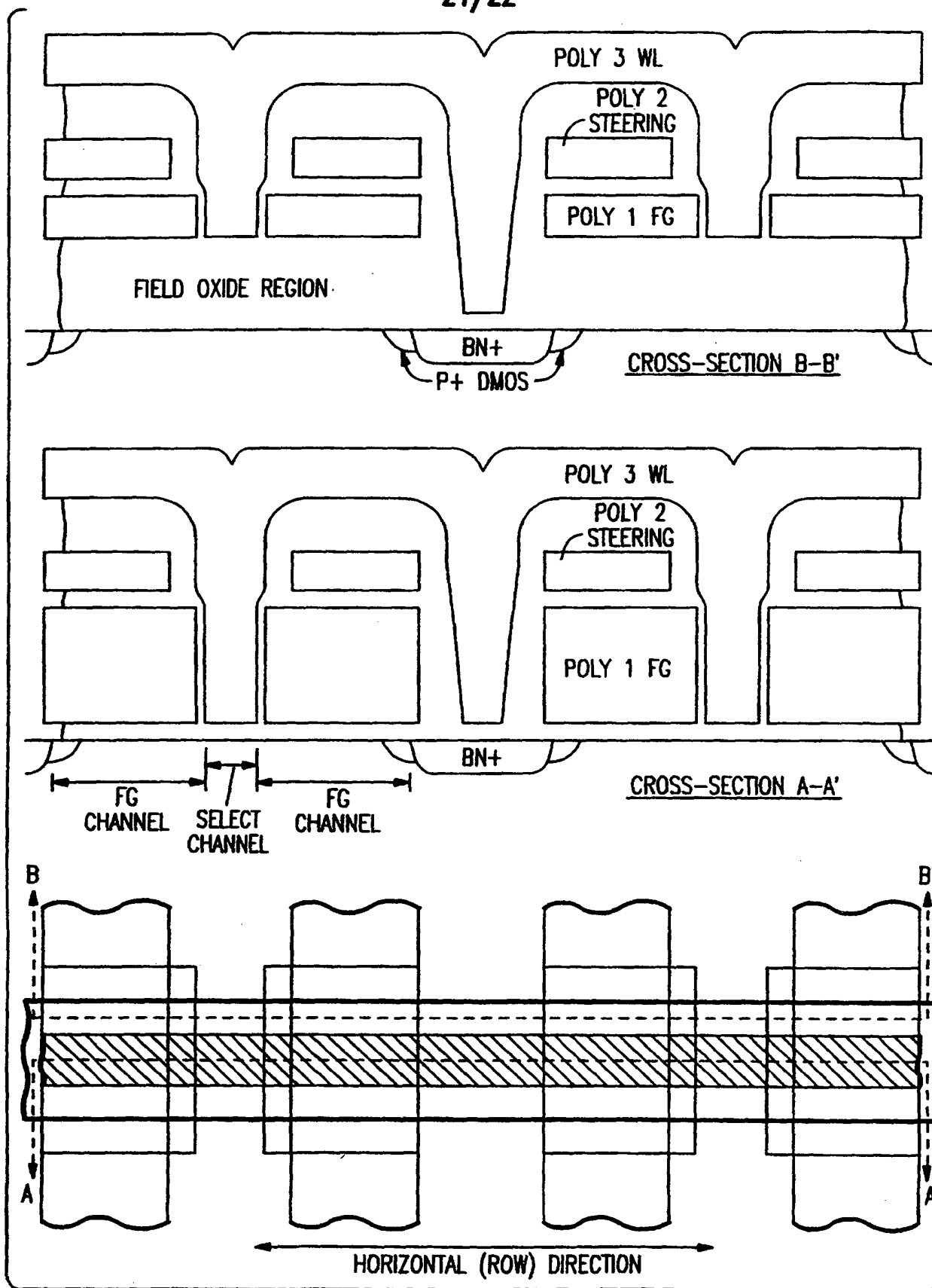
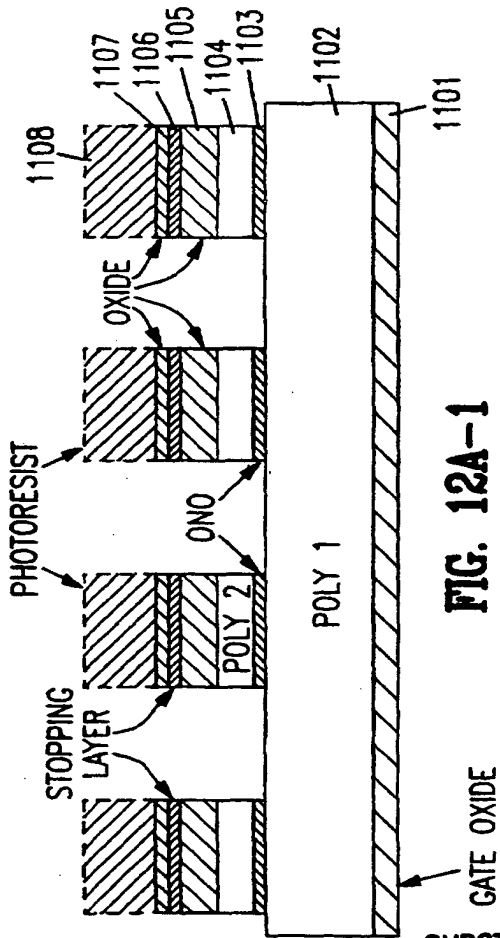
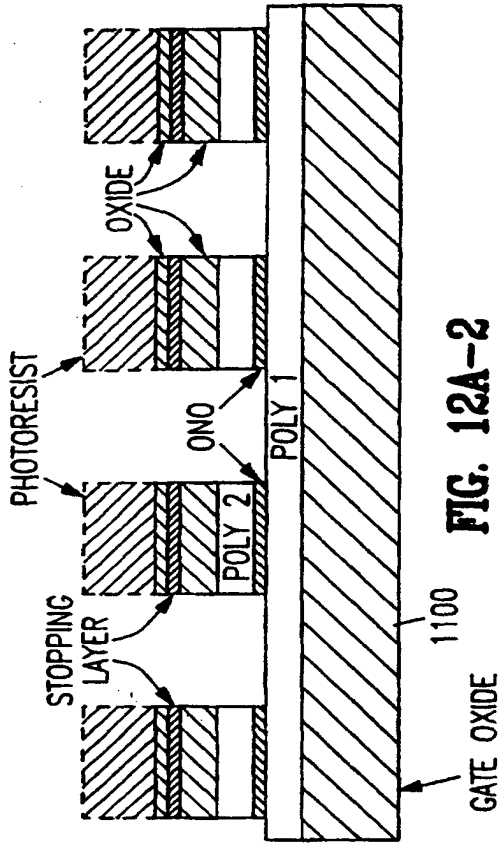
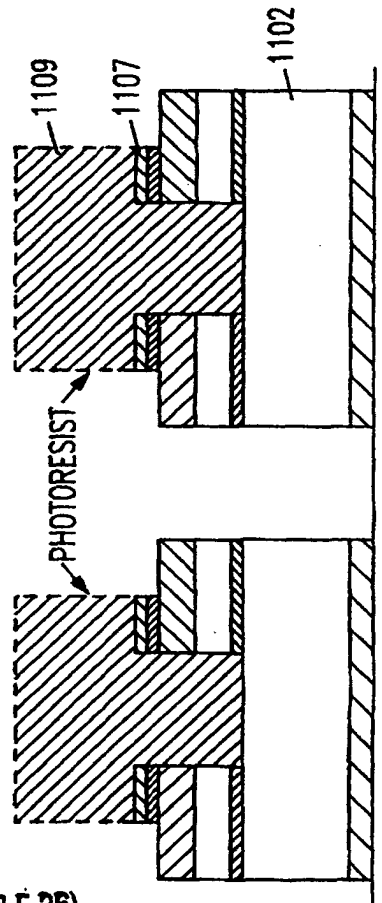
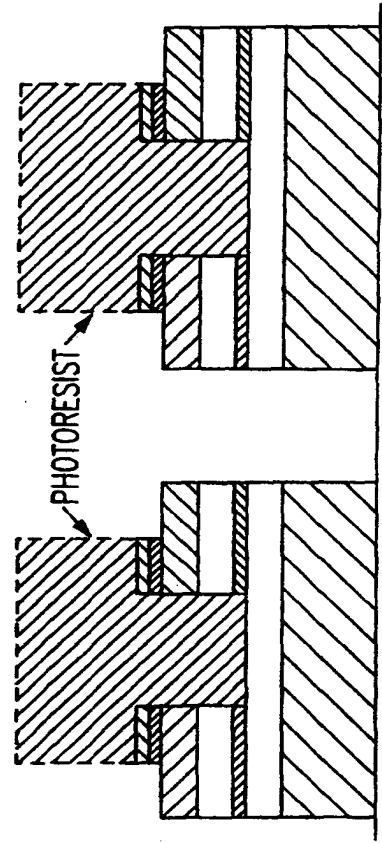


FIG. 11E

22/22



SUBSTITUTE SHEET (RULE 26)



INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/01388

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : G11C 11/00, 11/34; H01L 29/788

US CL : 365/149, 185.26: 257/314, 315, 316

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 365/149, 185.26: 257/314, 315, 316

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y, E	US 5,606,521A(Kuo et al.) 25 FEBRUARY 1997, see entire D document.	1

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	*T later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
*A document defining the general state of the art which is not considered to be of particular relevance	*X document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
*E earlier document published on or after the international filing date	*Y document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
*L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*Z document member of the same patent family
*O document referring to an oral disclosure, use, exhibition or other means	
*P document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

22 MAY 1997

Date of mailing of the international search report

09 JUN 1997

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile N (703) 305-3230

Authorized officer

CARL WHITEHEAD, JR.

Telephone N (703) 308-4940

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/01388

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

SEE ATTACHED

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
☒ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/US97/01388

2. Unity of Invention is lacking

1. This International Search Authority has found 2 inventions claimed in the International Application covered by the claims indicated below:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claim(s) 1-21 and 52 to 63, are drawn to a semiconductor device, class 257/315.

Group II, claim(s) 22 to 51, are drawn to a method of manufacturing, class 437/48.

and it considers that the International Application does not comply with the requirements of unity of invention (Rules 13.1, 13.2 and 13.3) for the reasons indicated below:

The inventions listed as Groups I and II do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: the semiconductor device could be produced by a different method or process. Further, the current sequence of manufacturing could be altered and still produce the same semiconductor device.